FOOD AND DRUG ADMINISTRATION

Immunology Devices Panel of the

Medical Devices Advisory Committee

September 19, 1997

Proceedings By:

CASET Associates, Ltd.
10201 Lee Highway #160
Fairfax, Virginia 22030
(703) 352-0091

PARTICIPANTS:

Panel Members:
Dr. Charles LaDoulis, Chairperson
Ms. Erika Ammirati
Dr. Henry Homburger
Dr. Glen Hortin
Dr. Wilbert Jordan
Dr. Mary Kemeny
Dr. Warner McCaskill-Stevens
Dr. Gustavo Reynoso
Dr. Sheila Taube

Mr. Peter Maxim, Executive Secretary, FDA

TABLE OF CONTENTS

	Page
Arlene Pinkos	6
Steven I. Gutman	21
Peter Maxim	31
Carolyn Jones, HIMA	55
Christopher Zalesky, Centocor, Inc.	60
Daniel Chan, AACC	74
Question One	81
Question Two	119
Question Three	133

P R O C E E D I N G S [10:03 p.m.]

DR. MAXIM: Good morning. I would like to welcome everybody to the Immunology Devices Panel meeting and get started with the meeting as close to schedule as we can. I know it there are several things that are going on this morning, so people will probably be filtering in for the first part of the morning.

To begin the meeting, let me read for the record the conflict of interest statement for this meeting of the Immunology Devices Panel.

The following announcement addresses conflict of interest issues associated with this meeting and is made part of the record to preclude even the appearance of an impropriety. To determine if any conflict existed, the Agency reviewed the submitted agenda and all financial interests reported by the committee participants.

The Conflict of Interest Statute prohibits special government employees from participating in matters that could affect their or their employer's financial interests. However, the Agency has determined that participation of certain members and consultants, the need for whose services outweighs the potential conflict of interest involved, is in the best interest of the government.

We would like to note for the record that the Agency took into consideration a matter concerning Dr. Glen Hortin. Dr. Hortin reported that in the past he conducted research relating to today's discussion. However, he received no personal compensation. Since this is a past involvement with no continuing financial interest and the issues before the panel are general in nature, the Agency has determined that he may participate in the Panel's deliberation.

In the event that the discussions involve any other products or firms not already on the agenda for which the FDA participant has a financial interest, the participants should excuse themselves from such involvement and their exclusion will be noted for the record.

With respect to all other participants, we ask in the interest of fairness that all persons making statements or presentation disclose any current or previous financial involvement with any firm, whose products they may wish to comment upon.

At this time, we have added several new members to the Panel. I would like to take the opportunity to go around the table and introduce all the members of the Panel, including the new ones.

We have on the side opposite from me, Ms. Erica Ammirati, who joins us as a first meeting as a industry representative and we are very pleased to have her on board and joining the Panel for the session.

We have Dr. Wilbert Jordan, who has served as the consumer representative to the Panel and we welcome him back for this meeting.

Also joining the Panel is Dr. Mary Kemeny. Dr. Warner McCaskill-Stevens has been a member of the Panel for the past couple of years.

Dr. Sheila Taube joins as a voting member for this session. We have Dr. Henry Homburger joining the Panel for the first time as a new voting member.

Dr. Reynoso, who is back with us -- we welcome you back. Dr. Glen Hortin is a new member joining the Panel for the first time today. And Dr. Charles LaDoulis, who has recently been appointed as chairman, chairperson, for the Panel for a term to commence immediately.

At this time, I would like to turn the remaining proceedings of the Panel meeting and Panel discussion over to Dr. LaDoulis.

DR. LA DOULIS: Good morning and welcome, everyone. I think the first representations are going to be

the FDA presentations on product development protocols and the new 510(k) paradigm that Dr. Gutman will present and the follow-up on the first year.

We hope to have time to complete all these presentations by the 11:30 break. If it is necessary, because of questions that we will obviously solicit, we may need to extend that beyond the luncheon hour, but we will certainly allow and provide for a great deal of participation from the public and the audience, of those that are present.

You know, we will have an open public hearing, as well as during the open committee discussion.

I don't have any other comments, other than to mention what we are seeking to accomplish here is to have as thorough a review of the current status of reclassification, as well as the final introduction of the new protocols, new paradigms, which have been sought for by panel members in the past and consultation with the members of the Agency in order to facilitate and promote the expeditious review of applications.

I guess the only other comment is that it really is a satisfying experience to see the introduction finally of this product development protocol because a number of

years of experience have brought to light the difficulties in dealing with premarket applications by panel members, probably in different panels, not only the Immunology Devices Panel in that submissions have been criticized and had difficulties because of the design of the protocols, which did not meet expectations of the Agency or did not meet expectations of scientific advisory panels.

There were opinions expressed many times that if there were opportunities for either scientific advisory panel members or the Agency to engage with sponsors in the design of studies, that would have facilitated the effective design of studies to address the safety and efficacy of immunology devices and certainly other products and for the sponsors, for the public, as well as for those engaged in the regulatory activity, it would have been a much more satisfactory experience in many instances.

So, I think that these new initiatives probably have been prompted for a number of reasons, but I think that the outcome would be of benefit to everybody, the public, the agency and certainly all those that are participating in the process of regulatory advice and recommendations.

So, without any further comment, unless somebody else has any comments before we begin, we will just go into

the presentations.

Anyone who would like to make any comments about today's agenda?

[There was no response.]

I will turn the meeting over to FDA staff and the first presentation. Introduce yourself, please.

MS. PINKOS: Yes, my name is Arlene Pinkos.

Good morning. I would like to tell you about a new initiative that is being proposed as part of FDA's reengineering efforts, called the Product Development Protocol or PDP, as I will refer to it. I would like to emphasize that the development and implementation of the PDP alternative is an ongoing process, being updated every week.

I will provide a brief introduction to the elements and processes of PDP and will be happy to answer any questions that you have.

If you have any comments after you leave here today, you can contact Dr. Lilian Yin, who heads up the Center's PDP Reengineering Team or you can comment via the Web. I will give you that address at the end of the talk.

PDP is intended to provide an alternative pathway to market for companies developing Class III devices, which would otherwise be required to go through the premarket

approval or PMA process.

Actually, PDP is not a new idea. The statutory authority for PDP was originally granted as part of the Medical Device Amendments to the Federal Food, Drug and Cosmetic Act of 1976. However, this alternative process was not implemented at that time because it was considered potentially complex and there was a need to focus attention on implementing the other core provisions of the Medical Device Amendments.

The intent of implementing PDP is to reduce both the resources required by FDA to review Class III devices, as well as the total time to get one of these devices to market. However, I should stress that the requirements for safety and effectiveness will be no less stringent under PDP than they are for PMA. Only the ways to which these requirements are satisfied differ.

Here is a simplified description of the normal route a medical device takes in order to reach the market, beginning with early concept development and progressing through various study and review phases. On the right, I have noted where various FDA involvement falls within this process. The sponsor might obtain an Investigation Device Exemption or IDE prior to conducting feasibility studies.

This is pretty rare for IVDs, due to the limited risk to the patients usually involved in the studies. More commonly, the IVD manufacturer would supply a PMA after all the clinical studies have been completed.

As you can see, the FDA gets involved much earlier in the process with the PDP. This benefits both the sponsor and the FDA because the extensive interaction between the sponsor and the Agency, which takes place early in the product development, decreases the probability of surprises later on. These unexpected occurrences or findings usually slow down or prevent approval of the device for marketing. In other words, the process is more proactive rather than reactive.

I would now like to go over some of the elements of the process. First of all, what devices are candidates for PDP? Candidates for PDP are those devices, which would otherwise be subject to premarket approval or PMA. And what about panel involvement?

Advisory panel review is still a required part of the regulatory process, although now, your input will come at a much earlier stage in the product's development. Instead of reviewing the data and study conclusions, as you do now in the PMA process, you will be reviewing and

commenting on the protocol and proposed acceptance criteria.

What is actually contained in the PDP? The proposed PDP must include descriptions of the device and any anticipated changes, all preclinical and clinical protocols, manufacturing methods, facilities and controls and proposed labeling for the device. Acceptable performance limits for each phase of the testing are also identified.

What steps are involved in the PDP process. Before you are the different phases of PDP, each of which I will be discussing in a moment. First, there are presubmission interactions, which precede the sponsors proposal to the FDA to enter their device into the PDP process. The sponsor then submits their proposal and the Agency conducts a filing review.

If the proposal is accepted, the sponsor submits their full PDP and FDA reviews it. The sponsor then conducts preclinical testing, clinical testing and then files a notice of completion. The Agency reviews the notice of completion for adherence to the agreed-upon protocol at which point the PDP is declared complete and then the product can go to market.

PDP can be thought of as a criteria-based developmental plan. When both FDA and the sponsor reach

agreement on the terms of the protocol, you might think of it as a signed contract. When the protocol is completed and all the terms that have been agreed to are met, then the sponsor will obtain approval for marketing their device.

I will now go over each of the stages of the PDP process. First is the presubmission. At this point, the applicant should consult with FDA, as well as any other outside parties, to develop the proposed PDP. The development of a PDP will require early and extensive interaction with FDA and consultants to provide in adequate detail all of the required information.

The applicant then submits a summary outline of the proposed PDP, not the complete protocol. At this point, FDA conducts a filing review within 30 days. The Agency determines whether the proposed PDP appears to be an appropriate candidate for this alternative process. If it is determined that PDP is an appropriate route for the device, then the sponsor submits the complete PDP for review.

At this time, FDA performs a substantive review of the entire protocol proposed. The studies described to support the safety and effectiveness of the device and all acceptance criteria are reviewed for appropriateness. It is

during this stage that the panel is sought. By the end of the 120 day period, FDA must approve or disapprove the PDP.

Following approval of the PDP, the applicant conducts their preclinical studies and develops their analytical data as it was described in the PDP. The PDP might stipulate that this data be reported to FDA in a progress report or, if needed, that the clinical phase cannot begin until these studies are completed and reviewed.

The sponsor will then move on to their clinical studies. Again, the terms of the PDP may require that information from the clinical studies be reported to the FDA in stages in the form of a progress report.

As the company progresses toward completion of the clinical protocols, inspections for conformance to Good Manufacturing Practices or the new Quality Assurance regulations, as well as Bioresearch Monitoring regulations, will take place.

Let me take a moment to talk about changes that occur during the course of preclinical and clinical studies. We recognize that not everything laid out in the original PDP will work out as planned and that modifications of device design, testing protocols or manufacturing processes may be necessary. These changes will require notification

to FDA, with prior approval being necessary if the change is substantive. Substantive changes will be addressed by the FDA within 30 days.

As throughout the entire PDP process, interactive meetings and teleconferences will be used whenever possible to expedite decision making. This is viewed as a very important aspect to the process.

When all the trials have been completed, the applicant submits a notice of completion to FDA, which contains all of the study data. This must be reviewed by the FDA within 90 days.

During this review, the FDA will validate that the protocol has been followed and that results satisfy the success criteria identified in the PDP. If all of the terms of the contract have been met, FDA declares the PDP complete and the product may then go to market, just as if it had been approved under the PMA process.

To summarize the key elements and time frames of the process again, the approval or disapproval decision of the proposed Product Development Protocol will be made by FDA within 120 days of receipt of the PDP. When all the clinical trials are completed, the sponsor files their notice of completion and the Agency has up to 90 days to

declare that the protocol is either complete or not complete. If it is declared complete, then the device may go to market.

In conclusion, it is anticipated that PDP will work best, at least in the beginning, for Class III devices, which are not first of a kind, and those for which FDA guidances have been developed. However, it is intended that eventually PDP will be of great assistance to the rapid development of innovative devices because it should be less expensive than the conventional two-step investigation and premarket approval process.

I want to remind you that this process is still under development. The most up-to-date information pertaining to PDP can be found on the Web at the address shown before you.

I would also like to mention that there is a PDP workshop to be held on October 22, to inform interested members of industry of this initiative.

Are there any questions?

DR. KEMENY: After the presubmission in the summary outline, you said it is deemed whether it is appropriate or not. What would not be appropriate?

MS. PINKOS: At this point, I don't think they can

really identify any device that would not be appropriate. I mean, that is the most honest answer. I guess what you want to have in that summary outline is a feel that the protocol that they want to do is going to work, that you would know how to judge the acceptance criteria. You would know how what steps to set up to evaluate the safety and effectiveness of the device; also, maybe things like the intended us. I mean, if somebody wants to make a really wild intended use that the Agency doesn't feel could possibly be supported by any studies, that might be rejected as well.

DR. KEMENY: The way it is worded, it just sounds like if something is complete, then the product can go to market, but what if -- you know, it is complete, they have done everything they are supposed to do, but it hasn't shown what it is supposed to show?

MS. PINKOS: Well, that is part of the tricky part of writing the protocol. I mean, the protocol is going to include a proposal of tests that will demonstrate the safety and effectiveness of that device. Then the acceptance criteria that go along with those studies have to be identified as well.

I mean, the studies that are proposed have to be

in both the sponsor and the FDA's opinion going to support the safety and effectiveness of the device when it is all over.

DR. GUTMAN: Can I clarify that, that it is my understanding that the acceptive criteria are like thresholds of performance.

MS. PINKOS: Right.

DR. GUTMAN: So, if you had a product that was dealing -- it itself was fundamentally new or the biologic process it was trying to define was fundamentally new and you couldn't establish performance parameters that would be reasonable for establishing safety and effectiveness, you could argue against using this process, that you have to know enough about the disease to know what your acceptance criteria should be and then if the sponsor fails to meet the acceptance criteria, the PDP could theoretically be rejected because you didn't meet what is reasonable in order to diagnose or evaluate this particular process.

DR. REYNOSO: Are there any provisions for during the process for the proposed product to go back to the panel or is that not contemplated in the protocol?

MS. PINKOS: That is not envisioned at this time. I would imagine if some twist or turn took place that was

totally unanticipated and the final results were not -- didn't quite meet the acceptance criteria that were established ahead of time, it could maybe go back to the panel.

DR. TAUBE: So, the panel's real involvement is at a stage where we provide advice on whether or not what they are proposing is reasonable and -- because it comes before the preclinical and clinical studies. So, we are supposed to judge whether or not what they think will be acceptable studies will, in fact, prove their intended use --

MS. PINKOS: Yes, and that the thresholds of performance are appropriate.

DR. LA DOULIS: I think there is a concern that is raised by a couple of the questions and also that I would raise for myself in that it is possible that what might be deemed to be satisfactory results of the clinical trials, according to the designs of the protocol that were initiated, might be still subject to debate among members of an advisory panel, even if we had been involved in the process of approving the design of the protocols and the standards of performance.

I think it has come often to the panel that the data that appear as they are presented to satisfy certain

criteria have under further analysis and review indicated some flawed analysis or some questionable criteria, which have not been previously anticipated.

So, I think that the involvement of the panel, scientific advisory review in some way, in addition to the notice of completion, ought to be considered as part of the process.

MS. PINKOS: I will pass that through along to Dr. Yin.

DR. JORDAN: That was part of my question. The other concern I have in trying to run this through my mind, it appears that our greater responsibility lies in your office or in the FDA person who is initially interacting with a proposer, since it is that person, who has to really review and either approve, send back, et cetera, before it comes to us.

I mean, it sounds like this person now has to have a lot of expertise in a particular field. And I would --

MS. PINKOS: Probably no more than in the PMA process. The whole emphasis of energy and resources is now just pushed earlier. It is a different mind set to start -- you are now thinking, trying to think ahead and trying to anticipate how studies will fold together and prove the

final safety and effectiveness of the device. It is definitely a different mind set.

DR. TAUBE: I actually think it is a wonderful idea to get involved at a much earlier stage and to help guide the manufacturer so that it is not a contest of I thought, you know, my ideas were better than your ideas. You are working together to get a product to market.

The only concern that I have is that when you ask for advice from the panel members or from whomever on just a protocol, it isn't always possible to predict the kinds of things that may or may not work. I mean, we are involved at the Cancer Institute in protocol review, but -- for drug development, for instance, but there is a lot of interaction that goes on after a study is started because sometimes the protocol doesn't work as planned.

So, what is the Agency thinking about in terms of once the preclinical and clinical studies are begun in terms of interactions, both with the panel, as well as the manufacturer?

MS. PINKOS: If I understand your question, what is the panel's involvement through the process, again, the panel -- at this time, the panel involvement is foreseen to take place while the PDP is reviewed and assuming that

everything works as planned and everything happens as you would expect it to happen, that there shouldn't be a need for further involvement of the panel.

DR. TAUBE: The issue was if everything doesn't work out as expected.

MS. PINKOS: Well, that is why the process is still under development. We haven't crossed all the -- we have a few submissions that are currently in house within the center and it is a learning process for us right now.

I would also maybe like to just take a moment to say that there are financial advantages seen to the PDP, one of which being that clinical studies can be very, very expensive and if you have got particularly a small company that doesn't know whether they want to invest in the cost of a clinical study by sort of having a PDP done ahead of time -- and that is not to say that developing a protocol is not cheap or doesn't take resources, but at least that a protocol would be agreed upon before the bucks are spent, so to speak.

But I agree with you. It is difficult to anticipate all the changes and that is why we compare it to negotiating a contract. When you are negotiating for millions and millions or billions of dollars, it is

difficult and you really have to think about everything ahead of time before you sign on the dotted line and it is definitely going to be a challenge

DR. JORDAN: Well, again, the idea is exciting in that it gets involvement earlier and, hopefully, allows approval faster, which I endorse. But, remember, in the last, and I participated, we didn't approve something, trying to look at that particular situation where if we had been involved earlier, I am still uncomfortable with no ability to evaluate the results of the protocol to know what happened.

DR. MAXIM: I think the important thing I would add to what Arlene has said is that, first of all, this procedure is under development right now also. We are certainly looking for input to this.

We anticipate on the Agency level that this could be a very interactive process and even after this initial panel meeting, where we have worked with the company. We have developed their -- helped to develop their protocol, their acceptance criterion and bring everybody together to present to the panel and get your input on it.

There is still going to be an opportunity for progress reports, evaluations with the Agency to let us know

how things are going and how they are proceeding towards meeting their objectives and these particular criteria.

Obviously, I think, mid process if something were to come up and surprises were to come along as part of the ability of the company to complete this contract or complete this protocol, there would be an opportunity to, you know, come back to the panel, either through teleconferences or homework assignments or if it was critical enough to reconvene the panel to gather input and either modify the outcome measures or perhaps suggest that this become a PMA and that we look at it on the other end after we have had a chance to evaluate all the performance characteristics of the device.

MS. PINKOS: Yes, that has been identified, that at any point in the process if things start going askew, the terms of the contract might be up for renegotiation. So, if it doesn't happen as planned, there are some safety factors.

MS. AMMIRATI: Not to beat a dead horse, but to maybe reiterate that is listening to the comments and that, I don't think I have ever been involved in a clinical trial where everything goes exactly to the protocol. I don't think there has ever been one. It almost sounds like we are using FDA as you would in an IRB. There is an analogy there

that you would have to do an amendment or there would be then a discussion whether this is considered substantive enough to either convene as a panel or do a teleconference.

So, just based on what you are saying, Arlene, that there is -- it looks like there is a nice established reiterative process, where something you started and, gee, this didn't work right. Let's look at it. Maybe we have to change a few elements and then decide whether that is considered major or not. So, there is almost some analogies to things we see in an IRB establishment or situation.

MS. PINKOS: The Agency has actually developed a modifications document, which outlines the types of changes that are anticipated that might occur during the process and we have attempted to categorize them either as substantive or non-substantive, requiring either prior approval before the change is implemented or not, at least give some guidance to what manufacturers can expect when those changes occur.

DR. LA DOULIS: Any other questions or comments?

[There was no response.]

I think this is a very important and a very hopeful process. I think it will be of benefit not only to the Agency, certainly to many companies, large and small, in

streamlining and forecasting how they might invest their time and effort. And I think the public will be better served by having probably an expedited review by this outcome.

If there are not any other questions or comments, we will go to the next presentation. Thank you very much.

DR. GUTMAN: Hi. I am Steve Gutman. I am the director of the Division of Clinical Laboratory Devices.

FDA is proposing a risk-based restructuring or reengineering of our work load. A central features of this restructuring is an effort to redirect resources to high risk/high impact devices. A consequence of this will be decreased review of low risk/low impact devices and the great challenge is distinguishing between the two.

Central to the issue of risk assessment is the FDA classification system. This system identifies products. It is Class I, II or III, based on increasing risk of use and on the type of controls or information needed in place before marketing.

Arlene has reviewed with you one tool being piloted for handling Class III products, the PDP. I would like to discuss a three step parallel plan for improving our handling of Class I and Class II products. This plan is

commonly referred to at FDA as the new 510(k) paradigm

Step 1 is an effort to revisit the basic classification of the menu of in vitro diagnostic devices that we regulate. The intent of this revisit is to update a system, which is increasing showing its age. I would remind you that the classification system that we currently use was developed in the late 1970s and early 1980s. There has been a lot of water under the bridge since then.

And while the classification system does hold its own, it is showing its age. Although we plan to look at all types of products, particular scrutiny will be directed at those currently designated as Class I. The intention is to identify low risk products with well-established technologies for the purpose of exempting these devices from the requirement of premarket review.

Class I exempt products are not a new part of the FDA regulatory process. Since 1976, for the entire office, a total of 573 devices have been placed in this category. Class I exempt products are also not entirely outside of FDA oversight. They continue to be subject to registration and listing. Generally, they continue to require compliance for the sponsor to Good Manufacturing Practices and they continue to be subject to postmarket reporting of harmful

events.

The division currently regulates over a hundred devices, which fall in the Class I category. Based on assessment of risk and novelty, we hope to exempt some of these products, make them Class I exempt, while up-classifying others to Class II levels of oversight. It is worth noting that DCLD is choosing products for exemption carefully. We do not plan to consider the products that are the subject of discussion today, tumor markers for this enterprise and that we will be seeking broad input on our choices in the near future, so that members of our panels, the industry and professional groups will have an opportunity to see what we have cooked up and to comment.

Step 2 is an effort to use enhanced manufacturing requirements as a surrogate for premarket review in a selected set of submissions. As of June 1st, 1997, manufacturers of Class II and certain Class I devices are required to follow what are called design control procedures, as part of their manufacturing processes. These procedures outline a systematic set of requirements and activities for the management of design and development of products.

These include the documentation of design inputs,

risk analysis, design outputs, test procedures, verification and validation procedures and formal design reviews. Manufacturers must, in short, ensure that design input requirements are appropriate so that the device will meet its intended use and the need of the user population.

The FDA hopes to exploit these new manufacturing requirements by using them as a basis for a new form of submission termed a Special 510(k). The Special 510(k) is intended as a mechanism specifically for reviewing product modifications. Product modifications in the production of IVDs are common and result from efforts by companies to improve or optimize their assays and/or to deal with changing supply conditions.

A guidance document was issued last year to help define when a modification is considered significant enough to require a new 510(k) submission. This document includes special sections and flow charts for in vitro diagnostic devices. It suggests that when intended use and technology don't change, as long as performance does not change in a statistically significant way and labeling remains the same, new 510(k) submissions are not required. The company is required to document the changes and maintain them on file, but as long as there is no statistically significant

performance change or labeling change, they are not required to submit a new submission.

When performance does change in a statistically significant way and as a result it is appropriate to change labeling, new submissions are required.

In the 510(k) paradigm, this new submission could take the form of a Special 510(k). This submission would be reserved for modification of devices, as noted earlier, not affecting intended use or device technology. Sponsors would declare conformity to design control as a substitute for conventional review.

Details of how this will be implemented and what controls on the controls will be put into place are the subject of ongoing exploration, but the impact, if successful, would be decreased oversight of changes we would view as relatively low risk in well-established devices. I might add that we estimate that perhaps 10 to 20 percent of our work falls in this area of modification. So, we are talking about a significant reduction in work load.

The third and final step in the 510(k) paradigm is a submission called an abbreviated 510(k). The abbreviated 510(k) represents an effort to introduce the use of standards as a component of the review program in a more

structured and formal manner than has historically been the case. The bottom line in this reengineering effort is that FDA will identify standards, which can be used in place of part or perhaps in some cases all of the review process for a selected set of devices.

Sponsors will follow those standards, summarize the results drawn from the standards in their submission and, as appropriate, explain how device risks are addressed by conformity with the standards. Again, the mechanics for this program are still being worked out, but a logical consequence is the need for FDA to work with industry and professional groups in improving the numbers and ensuring appropriate quality and, frankly, ensuring usability of existing standards.

My favorite quote from the industry we regulate is that dealing with regulators is like trying to herd cats and I might suggest that some of my most unkind co-workers would suggest that dealing with some companies is like trying to herd trees. That being what it may, FDA is clearly working on a new meow. It is our hope we can do business in a new and more modern way without compromising and perhaps actually improving the overall quality of the science and the oversight we provide the medical community.

I will be happy to entertain questions. I realize this is a dazzling array of regulatory nuances that I have thrown at your feet.

DR. JORDAN: Is it permissible for you to give an example of a Class I, a Class I exempt, a Class II and why the Class I exempt is not a Class II?

DR. GUTMAN: Yes. The Class I exempts historically were general purpose items that actually did not have performance characteristics. So, collection cups, beakers, general purpose culture media were considered Class I exempt originally and there are -- general purpose culture media, some of the simpler transport devices were considered Class I.

In the revisit to Class I exempt, we are crossing the line. Before, the general rubric was that if a device had performance characteristics, we were not willing to consider exempting it. And in the proposal that is on the table and that will be publicly outlined sometime in the near future, we are taking simpler devices and more established devices that do have performance characteristics and based on the fact that it is non-novel technology, that they are products that have been around a long time and that they tend to be products that by themselves don't make

diagnosis, but make diagnosis in the company of friends.

We are going to be proposing that some analytical parameters actually be brought down and it will be fun to see -- my prediction is that the industry will think that we have been too conservative and that the professional groups will think we have been too liberal. But my guesses aren't always right. I am not Jeanne Dixon.

DR. REYNOSO: I realize that the use of a standard is a new and ongoing process, but could you elaborate for a few minutes on that, perhaps a couple of examples?

DR. GUTMAN: The standards that we know and love best, the standards that we have had actually the most professional interaction and input within are the NCCLS standards. Those standards are actually variable, but, in general, a wonderful set of standards. They are not defined specifically toward manufacturer submissions. They tend to be more laboratory oriented, but the protocols they develop and the questions they ask and the data sets they generate apply to wide ranges of our product line.

The target branch for the standards activity and the abbreviated 510(k) is the Chemistry Branch, where we have a wide variety of simple quantitative methods, which could be plugged into various standard evaluative

techniques. They tend to raise very well -- you look at analytes(?) like the sodium in creatinine and BUN and all of your old friends -- by now, probably most people understand the clinical and technical nuances that are involved, at least in the old technologies, and they flow immediately into the notion that you can standardize a package and make it really easy.

There are other standards as well. Some of the professional groups write monographs or have their textbooks, some chapters that are sort of surrogate standards. The international groups, the ISO and the DIN(?) and some of the international groups have standards as well. They don't always lend themselves quite as nicely to our review process.

And I think the ultimate in standards would be if we could interact with industry in developing our own sort of tailor-made standards that might borrow liberally from NCCLS, but make them a little bit more parochial for the needs of our submissions. That might be a nice logical long term outcome.

One of the purposes of freeing up time from doing less review of low risk products is so that we can do better general direction in terms of standards and better review of

high risk products.

DR. TAUBE: In your discussion of design controls, you said ensure that input requirements are appropriate to intended use and user needs. Can you define "input requirements"?

DR. GUTMAN: Well, I am not an expert in design controls and we have actually looked at the concept of design control -- historical context, design controls, and the entire new GMP process is based on more of a quality management system and it borrows rather liberally from the system that is being proposed or I guess being used in Europe, the ISO 9000 and 9000 series.

I actually cannot comment. I really am not comfortable with the semantics and with the particulars and we have recently talked to the people in the GMP area, asking them to start looking at guidance that we could understand how all of those things flow into an IVD, in particular. At least a preliminary response we got back is that they do seem to know what is going on and they do seem to provide guidance and one of our priorities is, in fact, to interact with our own GMP folks.

One of the areas that I personally am weakest on -- I am going to ask either Max or Kaiser if either of them

has more insight into the design control than I do. They may not, but that is not something I have great strength in. I have great interest in because if we are going to make this program work, we have to understand design controls. We have to be able to communicate it to people who are inspecting firms as a surrogate for our review and we have to be comfortable for the link between the two.

So, it is an important question but it is one I am -- it is beyond my intellectual realm at this point. It won't be in the future, I assure you. I have asked actually HIMA to put together a working group to interact with us and help us understand maybe based on the European model.

Max or Kaiser, you guys both have been somewhat involved. Do either of you have any wisdom?

DR. RABINOWITZ: Max Rabinowitz.

I don't work directly with the design controls, but I have attended a couple of workshops. It is my understanding that the input is the requirements for the device. What will the device do? And then the output is the finished product and the specifications. So, the user will have the specifications, will be able to verify that the product meets the specification, but the input is the manufacturer's concepts and what has to be achieved in the

process.

DR. GUTMAN: Any other questions?

Thank you.

DR. LA DOULIS: If there are no other comments, then we will -- I guess we will proceed to you, Peter, for a first year of tumor marker reclassification.

DR. MAXIM: Welcome to the panel and guests.

The purpose of this session primarily today is to give a one year update on tumor marker reclassification and to provide -- get some comment from the panel and from interested individuals in the audience on the special control that is being used to regulate these Class II tumor markers.

I would like to go back in time to December 1st of 1995, when this panel met to consider the reclassification petition for the tumor markers, that they be moved from Class III to Class II. There are enough new members of the panel today that I think I will give a few minutes update of the history of where the panel has been in the past to benefit them as we move on for the rest of the considerations in the afternoon session.

At that time, the Agency had a petition before them to reclassify those tumor markers that are used for

monitoring either recurrence of disease or response to therapy, those tumor markers, transitional devices, from Class III, which required a premarket approval into Class II, which could be regulated by the 510(k) route and through the use of special controls.

As it states at the bottom of the slide, the request was specifically and uniquely indicated for those tests that were being used for monitoring patients that had previously been diagnosed.

At that time, the number of tumor markers that had gone through the PMA process and had been approved are listed, the CEA, AFP -- there was one approval for a CA125, prostate specific antigen, cancer antigen 15-3, which, in fact, at that time was CA27.9, had just recently been approved for us, and VTA, which is a bladder tumor associated antigen, had recently been approved by the panel.

So, the numbers of PMAs and the numbers of products on the marketplace were relatively limited if you consider the fact that one of the first tumor markers was probably approved in 1983 or 1984 and the agency thought that this would also help allow these markers for monitoring to get into the marketplace in a quicker fashion, in a more expeditious fashion.

We felt that the basis for substantial equivalence for these markers could be established and dealt with if they were, in fact, moved to Class II. We would look actually at a comparison to a predicate device if one was available. We would look at the description of the literature base, the supporting information for that particular product, whether it was product specific, analyte specific and either called for or look at clinical trials that had support of that data. Analytical data requirements are listed and are characteristic for most of the things we looked at and that clinical data to support the sensitivity, specificity, the cutoff and the decision point for the markers and whether or not the marker could, in fact, monitor patient course, would be available to the Agency not only for some of the older markers that we had experience with but certainly with newer markers and we could evaluate review and call for this information as part of the 510(k) process.

Another significant aspect to dealing with these markers or any product for that matter as part of the PMA process is the administrative issues that are associated with the procedure itself. This slide summarizes the types of requirements for Class III products, compared to a Class

II product that would come through for a 510(k) of administrative issues, but by far and away the most time-saving elements for both the agency and for the manufacturer would be the preparation of the summary and safety and effectiveness data for PMA.

Often time, this is a multi-page document that is put together by the Agency and published, which allows the public to know essentially what we reviewed as part of the safety and effectiveness information for that particular product. For a 510(k), this can be a summary of information provided by the manufacturer and often does exceed one or two or three pages in the front of the submission or the manufacturer can select to provide a statement to the Agency that he will make that data available to any interested parties upon request and on final clearance of the 510(k).

This also now no longer calls for premarket GMP, inspections of the -- or premarket approval of their GMP inspections before a product can go to the marketplace and, therefore, that requirement indicates that the sponsor does not have to put together an entire manufacturing section for the submission as it comes to the Agency.

Of course, there are some other parts that are variable to both submissions and we look at the panel

review, which is not necessitated for a 510(k) submission, but can be called upon or called for.

The special controls that were suggested to be included for the tumor markers as they went to Class II was the guidance document written by both industry, professional associations and the Agency, put together as a group effort, was provided as part of the reclassification petition by Centacore(?), who brought the petition to the Agency and it was reviewed by the panel as part of the reclassification process.

We also called upon, as Dr. Gutman just mentioned, some of the controls and standards put forth by the National Committee of Clinical Laboratory Standards for method comparison, bias estimations, precision, et cetera, and those could also be used in the review of the submissions for Class II products.

The panel voted at that time that, in fact, these markers could be dealt with as Class II products. They voted unanimously that the Agency should move ahead and position down -- reclassify them in Class II. At that time, we then proceeded to move ahead towards that goal. One of the most memorable events that I have at that meeting was someone from the audience asked a question as to how long

this will take and I believe my estimate was 18 to 24 months. And as most of my friends from industry will remind me, as with any other guess as to how long something is going to take, I was wrong.

We, in fact, were able to reclassify these by notice of approval to the petitioner, Centacore, rather than go through formal rule making. So, that 18 month period was reduced to nine months and on September 19th, 1996, these products did -- were reclassified from Class III to Class II.

Today, we need to go back over where we have been over the period of the last year. As I have mentioned, the products were reclassified. The official order was sent to Centacore on September 19th, which was a Friday. On Monday, we had the first submission in my office that deal with these markers. And since then, we have had a total of 30 510(k)s come to the Agency for tumor markers.

The distribution of these tumor markers is presented on this slide. Two of them represent colorectal type markers or analytes, three testicular, 12 prostate markers, four ovarian, six breast markers and three bladder cancer markers.

Of the 30 submissions that have come to us, the 30

received in the first year, we have determined 13 of these markers to be substantially equivalent and have granted clearance for them to the requesting company. Two of these 13, in all fairness, were PMA conversions. These were companies that got caught and had PMAs in house when the tumor markers were reclassified. We got together with those two companies and discussed the options available to them and they opted to withdraw the PMA and bring the same data package and the same types of information back to us as 510(k)s. So, they were cleared in probably a relatively shorter period of time than if we continue on processing those PMAs.

One submission has been withdrawn since coming into us and two submissions came in and we refused to accept them because of both administrative and data requirements of the submission. They just weren't were in an acceptable form that we felt that we could go ahead and review them for the company.

Of the 13 that we have determined to be substantially equivalent, we have granted clearances to. The FDA review time for those 13 in days ranged from 29 days to 216 days, with an average of 124.8. Even on average of some of these markers, it exceeds the 90 day review time for

a 510(k), but I think most of industry would agree, it is somewhat faster than having a PMA being reviewed and processed by the ODE.

The number cleared again was 13. The number cleared on the first cycle, and this is within the first 90 day review cycle, was five and that comes out 38.5 percent. The review time range, again, is the 29 day to 90 days allowed in the -- the 90 days that is allowed as part of the first cycle.

The number cleared in two or more cycles, obviously, is the remainder or eight. The FDA review time ranges from 113 days to 216 days on those submissions with an average of 157 days that we were evaluating the submission in house.

The non-FDA time, which is the difference between the day it came in and the day it left the Agency, minus the FDA time, was 24 to 134 days. So, again, this is a popular program and industry is coming back and supplying the information, answering the questions and the data that we need to grant clearances for this.

So, overall, I feel that the program has been fairly successful. We have put -- we have been able to clear quite a -- 13, in fact, of these markers within a

period of a year, which, trust me, exceeds any performance that we had in the past if we were dealing with these as PMAs or PMA supplements.

Now, to go over key components of the guidance documents available as a special control and I will go through these briefly because what we want to do is focus on the questions that we want to come back and evaluate this afternoon and have the panel discuss and get comments from the public if they care to comment on the guidance documents. I will try to move through these fairly quickly now.

The guidance document focuses on, obviously, the administrative requirements common to all 510(k)s and anyone who does the submissions are familiar with these. You need a summary or the statement of availability of the safety and effectiveness data for the submission. There is a truthful and accurate statement that has to be filled out, the indications for use statement that the Agency signs off on, information addressing the use of the special controls and the table of contents and appropriate pagination.

The next aspect that we go to then would be the analytical data, the non-clinical studies. We look when appropriate, when necessary, at reagent characterization,

the assay specificity and interfering substances that have to be addressed or dealt with as part of the particular marker that you are evaluating or submitting.

Non-clinical studies continue with the performance characteristics of the device. We look at analytical sensitivity, the smallest amount, the minimal detectible concentration, whatever you want to call it. This is essentially the smallest amount of the analyte that we detected with the device, reproducibility. We look at the linear range of the assay over the working range for the product. We look at precision, the precision data supporting it, multiple levels, and if there is any information that can be supplied regarding a high dose hook effect for this particular product; other parameters that need to be evaluated and data supplied, if appropriate; storage conditions; whether or not you are dealing with plasma compared to serum, serum plasma comparisons to show that there is no plasma interferences. You get the same test results. And, of course, if you are going in other matrices, the fact that you can measure the analyte in that matrix with the appropriate accuracy and sensitivity.

As part of the 510(k) process and if an issue arises, we may ask for information regarding stability data

or simply require whether or not the manufacturer has sufficient stability data on hand to support the dating of the product.

Another aspect is comparison studies. Again, you can compare your product's performance on a series of samples to a legally-marketed device. Here we look at linearity. We may calculate relative sensitivity and specificity off of the samples that were used in that particular study, linear regression analysis and at times a resolution of discrepant results between the two devices that may become an issue in the review.

All of these are addressed in much more detail in the guidance document.

The predicate device that we are referring to is the -- or the comparison device, depending on how you want to look at -- the predicate device legally is a product that was marketed prior to May 28th, 1976. If this device was in the marketplace at the time of the passage of the medical device law at that time, then new manufacturers who would come to the marketplace via the 510(k) procedure, simply saying the performance of their device was substantially equivalent to something that was already in the marketplace and didn't have to follow new premarket approval

applications and gather the clinical information to support safety and effectiveness at that time.

It can also be a device that has been reclassified from Class III to Class II or to I, which these tumor markers were now falling into. And since 1976 and in the 1990 Medical Device Act, we have allowed this to also include a device with a valid 510(k) on a product that has been seen and approved by the agency in the past.

So, this is the predicate device upon which you need to evaluate your device for a determination of substantial equivalence, The Agency looks at substantial equivalence to this predicate device as being a function of its intended use and this is in a flow diagram. This is in front of the 510(k) jacket.

It is a function of the intended use, the technological characteristics of the product and certainly the performance characteristics of the product, not only analytical performance but clinical performance also, so that the determination of substantial equivalence is not truly simply on any one of these assays. It is not basically an analytical aspect but is analytical, clinical with regards to the intent of use. It is all these woven together into a substantial equivalence determination.

We talked to comparison of two predicate devices. This is some excerpts that I have taken from the literature showing, for instance, a breast cancer marker that has been being evaluated by two different procedures and we can see that in some cases the comparison or the agreement between the two assays is a very good, close and tight fit. In this case, you would certainly expect that as being the same marker, basically the same type of analytes, but it is simply by two different procedures.

Often times when you start looking at comparisons between two different markers when your predicate device or your comparison device is not precisely the same marker or the same analyte that you have -- and I want to emphasize at this point that this is an option. You don't, for instance, have to be measuring analyte A and find a predicate device or a comparison device that measures analyte A. With the appropriate supporting clinical information, you can compare yourself, your performance and your basis, your substantial equivalence to another marker, perhaps analyte B.

In this case, in that situation we have analyte A as a predicate device or a legally marketed device and comparison with another product or another analyte. And, again, I believe, this is in the literature also for these

two markers.

Here, again, with the slope we see a fairly decent correlation between the two analytes. The scatter is obviously at very high levels of the analyte and certainly out of the medical decision range. The cutoffs for products of this type are generally in the range of 30 to 40 units per ml.

Still, there is some scatter at this point and the Agency would generally ask for these points to be split out and take a look at this type of performance of the product at the medical decision point on a larger scale, so that we would stretch the scale up from zero to fifty units, rather than the current axes and see how well it performs right around the decision point.

We can often look also at clinical information or clinical data in support of a 510(k) submission to establish the performance to determine whether or not the product is, in fact, substantially equivalent to one that is already on the market. Again, we also can ask for fairly significant studies for clinical information to support premarket approvals in order to -- for an assurance that the device is safe and effective under conditions. So, clinical data requirements are not limited to the premarket approval

applications, but they can also extend to 510(k)s. Depending upon how far we remove ourselves from the device where we are looking at substantial equivalence, as you get farther away from not being exactly the same analyte as determining to say a marker for lung cancer substantially equivalent to CEA, which is used for monitoring recurrence of colorectal cancer patients, that type of a distancing from the other two submissions of the two 510(k)s would lead us into what is called a tier 3 510(k) and certainly ramp up the importance of clinical information to support the use of the proposed device.

Again, the clinical studies -- and these are outlined in the special control in the guidance document, which is available -- would call for a protocol on how you are going to conduct the study, who you are going to enter into it, the study sites that are going to be involved and we list the investigators.

Again, the clinical studies could rely upon gold standards if they are available and there are very few of these, reference methods, reference procedures. Often times we refer back to clinical status, what is actually happening with the patient with regards to the in vitro diagnostic or the device that you are looking at and then, again, the

comparison methods are important, can be an important aspect of the clinical studies also.

We look at clinical sensitivity then, how well the sensitivity, the ability of the product to detect the disease in patients that actually, in fact, have the disease or have the investigator parameter we are looking at, the specificity, clinical specificity of the product. We look at the validation and challenge of the cutoff. If the manufacturer has established what the cutoff is for the device, we take a good look at whether or not that is being validated as part of the clinical studies and we certainly look at the effectiveness in the population of the intended use, which is the purpose of the study in the first place.

Again, the clinical evaluations, there is some debate as to how important they are and how far we go, where they are necessary and how extensive they are. Recently, with two products that have come through the Agency, the first marker, and, again, when you look at breast markers -- and I guess this is all public information. It has been published in both the SS and ED product labeling and has been reported in the literature, everything from the medical journals to the gray sheet.

We have the performance of this particular marker

in a prospective clinical trial that this panel reviewed a couple of years ago. There was 166 patients entered in to the study and of those 166, 26 recurred. The device was able to detect recurrence in 15 of those 26 and then that had a clinical sensitivity of 50 -- whatever that is -- 52 percent and a relatively high specificity.

This product was on the market then, was available as a part of reclassification to be used as a predicate device. We asked, therefore, then that the company that came in with the next marker to also conduct a clinical investigation of this type to determine their performance in the intended use population, which is the ability to check recurrence in stage 2 and 3 breast cancer patients.

We can see when they use their marker in a similar type of -- in a similar -- I am not going to say the studies were conducted and they obviously weren't identical -- in a similar population with a closely related number of patients, that the sensitivity calculated on the device was 38 percent with specificity -- again, a fairly high specificity of 96.2 percent. The company, however, had available to them as part of their clinical studies the comparative device.

When they looked at the exact same patient

populations with this other product, they found, using their cutoff, using separate cutoffs for each device, they found that the sensitivity of the product in that identical population was 41.3 percent with a 92.4 percent sensitivity.

Some would argue that this is redundant. This is not necessary, that the Agency doesn't have to see this type of a data. We feel, however, that I think this provides a significant amount of information to the users and the medical community, the laboratory and the medical community, that you have these products, that under these particular conditions, they perform -- this added information, they perform individually in a particular manner and they could be quite comparable when used on an identical set of patients.

They also provided in the labeling was the relative sensitivity and specificity of the devices were comparing one to the other at their appropriate cutoffs. So, again, with the second device going to the marketplace, this information is available in the package insert of in the product labeling for the users that manage to see how these two products perform or how this product performs under these conditions.

We are going to come back this afternoon and this

is a -- we are going to have the panel primarily address this question. It is one that is the Agency has been going around and around with. We have been talking with our colleagues at industry about this type of information. We get input from the professional associations and today it is your turn to deal with it.

What is an old versus a new marker? When does a new marker become old and how do we feel -- what are your opinions on the so-called "me, too" devices?

You have seen some of the information in the data that we have had. If the Agency has looked at and examined 15 CEA assays and a new one comes to the door, a manufacturer wants to put it on his or her instrument, that probably does classify as a "me, too" device. We have had considerable experience with it.

What is your input as far as actually looking at the data requirements and particularly the clinical performance of that product before you are comfortable taking it into your laboratory and working on it, before you are comfortable getting test results off that particular product?

Do you consider the second product to enter the pipeline, to enter the FDA pipeline, a "me, too" product,

too? Has the new marker for which there is one clear or one approved use become old because it has been cleared by the Agency? Does that mean the data requirements are reduced, the amount of clinical information we look at has to be -- should be reduced or can be reduced? And if so, by how much?

Does the product become old after third reiteration, fourth or tenth? This type of information is what we are looking at.

So, that is one question that we are going to couch in a different format this afternoon for your benefits, but when is a new tumor marker no longer new? And we have already alluded to many of these types of questions or I just alluded to many of these that we will be looking at and evaluating here this afternoon.

At what point is there no longer a need to review clinical performance in the target population? Is it sufficient to say "yes," we do compare. We have this linear regression analysis or side-by-side comparison with another marker. We show -- and we can show how the antibody system interacts with normal benigns, the cancer population in question, other cancer populations, so we can get an idea of the specificity and distribution of the marker or do we have

to go to seeing side-by-side comparisons and the ability of PSA to detect recurrence in men who have undergone prostatectomy or are under prior treatment for prostate cancer?

So, that is the gist, the nature of this type of question. We will be looking at that again this afternoon and hand in hand probably, just another rephrasing of this is at what point is analytical data and comparison to a predicate, side-by-side comparison, sufficient for product clearance?

That is basically what I wanted to introduce to you this morning. I think the hour is coming close to the time when we were scheduled to break for lunch. We will come back this afternoon and in part of the agenda we will hear from three representatives from professional associations and industry, who want to express their views. We will present scenarios or questions again based on this type of -- those questions we went after this afternoon and ask for your discussion of that this afternoon and how you feel about this.

As Dr. LaDoulis mentioned before, perhaps get additional comment from the public after the panel has had a chance to comment and discuss these issues. It is very

important that we look at this again. I think that these questions, at least, in one form or another have come up before the panel on two or three different occasions. They certainly came up as part of the original panel meeting for reclassification and I think they warrant a large amount of discussion, in-depth discussion this afternoon so we can take a real good look at these products as we move ahead.

MS. AMMIRATI: I just want to make sure that -- obviously, it made sense, but of the 30 you have seen in the last year and 18 have been dealt with one way or the other, there is 12 in the queue in various stages of being reviewed?

DR. MAXIM: Yes. I don't deal with numbers for obvious reasons. Yes. Either on the queue, on hold, requesting additional information of some nature.

I would like to make a comment at this time also that when this happened, as I said, we had one submission two days -- the Monday after reclassification went through and it was a short trickle until like December or January and then we had a fairly decent number of submissions.

I want to take this opportunity to express my appreciation to my staff also because this has been a phenomenal undertaking for them in conjunction with

everything else that goes on within the division. The first batch, I think the first 15 to 20 of these -- well, practically all these submissions have been dealt with within 70 days on the first cycle. We look at them. We review them. We decide if we need additional information, what information we need. We either call the manufacturer or send him a letter and put the submission on hold until we get that additional information.

We have done that almost to a hundred percent within 70 days and then, obviously, when the submission comes back in again it goes in the queue but it is evaluated very quickly afterwards. You know, I want to take this opportunity to express my thanks to them because they have done a phenomenal job over the period of the last year dealing with these and handling these submissions.

DR. LA DOULIS: Are there any other questions about what the aims of the discussion will be in the agenda for the afternoon? Will there be some members of the staff that might address some of the statistical issues and criteria in the guidance documents with regard to any points they might want to make? For example, this issue about predictive values and sensitivity and specificity, which have been expressed as percentages and reviewed by panels, I

think the issue might be addressed in the future as to whether or not there should be some standards adopted as to what are significant and/or acceptable levels of sensitivity.

DR. MAXIM: We can certainly address that. I will have to check and see if there is anybody from our statistics staff here this afternoon that could speak to that directly. Otherwise, perhaps, one of the reviewers or medical officers might want to deal with that issue this afternoon.

Adopting standards towards the review of this is certainly one criteria that we are looking at. That can be part of the discussion and even beyond that as to whether or not we need to do this on a case-by-case basis after a certain point in time. We will see if we can't have some discussion on that this afternoon.

DR. LA DOULIS: I don't know whether or not the issue has been addressed as to whether or not in addition to the usual descriptive parameters, whether the variance of the scatter, which is a statistic that is always asked for in correlations, whether that is something that is a parameter, which could be included in the standard evaluation, in addition to the straightforward descriptive

statistics because I think that has been the subject of debate and discussion in panels as to what the variations mean and whether or not an assumption of a linearity is appropriate in evaluating a new set of data.

DR. MAXIM: This is something we are certainly looking at because if we minimize, reduce or move towards a situation where we are taking less clinical data on some of these submissions, I would think that we would have to look at the analytical parameters quite closely. Several issues that come up that we are addressing on a day-by-day basis is exactly what. What type of regression analysis do you use? Which is more appropriate? Which data points, what range of the curve should you really be looking at?

For instance, the PSA for monitoring, although the four cutoff is widely accepted and great, that is a detection cutoff and what we really need to do is see the performance of that assay between, say, zero and 1 nanogram per ml and how well it is performing in that level, if it is being used to detect and monitor these patients for recurrence.

There has been some talk within the Agency and at some of these difference sessions that we have of looking at bias plots on side-by-side comparisons, rather than the

regression analysis for the markers where you could see the differences between the two assays or measure the differences between the assays more closely.

We have, I think, both on the panel and in the audience some excellent laboratorians, who have quite a bit of experience with tumor markers, who may want to talk to that or address it also.

DR. LA DOULIS: Okay. I don't know if there are any other questions. Since we are right at 11:30 and if there are not, then we will adjourn for lunch and reconvene at the scheduled time, at 12:30 here.

[Whereupon, at 11:30 a.m., the meeting was recessed, to reconvene at 12:30 p.m., the same day, Friday, September 19, 1997.]

P R O C E E D I N G S

Agenda Item: Open Public Hearing

DR. LADOULIS: As we go into this afternoon session, I will remind the panel members to make sure you are speaking into the microphone. You are having problems being heard on the other side of the room. And please identify yourself for the reporters before making comments. It will help make their jobs a little bit easier, also.

Okay, let's start. Welcome back, those of you who were here this morning, and any others who have just registered. This being the open public hearing with presentations scheduled by those who have requested to make comments from the Health Industry Manufacturers Association, Centocor, and American Association of Clinical Chemistry, unless there is any other business, housekeeping issues, then we can begin right away with our first scheduled presenter, which is going to be from Ms. Carolyn Jones, on behalf of the Health Industry Manufacturers Association. Carolyn?

Agenda Item: Health Industry Manufacturers Association Presentation

MS. JONES: Good afternoon. I am Carolyn Jones from the Health Industry Manufacturers Association. I would

like to thank FDA for allowing me to discuss the issues with this panel today.

HIMA is a trade association, a Washington, D.C. trade association. It is the largest medical device association in the world. We represent over 800 manufacturers of medical devices, diagnostic products, and medical information systems.

Our members manufacture over 90% of the $55 billion healthcare industry technology products. For many of our members that manufacture in vitro diagnostic products, the regulation of tumor-associated antigens, or tumor markers, is an important issue.

Regarding the regulation of tumor markers, we believe that FDA should be commended for the comparative speed with which it reclassified tumor markers. Historically, this process has taken from two to eight years. For tumor markers, only nine months elapsed between the Advisory Panel Meeting and the actual reclassification.

The speed and positive outcome of this process again demonstrates the value that results when FDA and industry work together. Although not speedy by comparison to other 510(k) reviews, the tumor markers which have been cleared, have been cleared with comparative speed.

Under the PMA process, sponsors could expect new tumor markers to be under review for one year or longer. As promised, reclassification has significantly reduced the administrative burden for sponsors and FDA alike. This will pay continuing dividends to both.

The central element in reclassification was the guidance document, or special control, which is the subject of today's discussions. It is essential to the regulatory process that manufacturers can rely on FDA guidance as acceptable to FDA. Unfortunately, FDA has not followed its own guidance document. Without providing any basis for the requirement, FDA has persisted in requiring studies in the target population for me too markers, even though these studies are not required under the guidance document.

A primary goal of reclassification is to eliminate the need to reprove what has already been established by the first of a kind marker. Once utility has been demonstrated by the first marker, there is no clear scientific or regulatory need to repeat the studies in the target population for another test, for the same analyte in the same matrix and with the same intended use and indications for use.

FDA's goal in requesting studies in the target

population is not clear. The purpose of the 510(k) process is not to demonstrate safety and effectiveness, but to demonstrate substantial equivalence to another legally marketed device; that product labeling is accurate and that instructions for use are adequate. The extensive analytical comparisons described in Section 6A of the guidance are sufficient to demonstrate substantial equivalence and to support performance claims in the product labeling.

Factors unique to tumor marker studies generally limit the ability to make study to study comparisons, thus these studies provide neither FDA nor the user with useful information regarding comparative product performance.

In addition to the questionable benefit of these studies in target patient populations, there are practical limitations impacting the utility of such studies for me too markers, and these limitations include selection bias resulting from the use of other similar approved or unapproved tests; the small study size; the protracted study duration; the impact of patient treatment and therapy decisions.

Even with unlimited time and financial resources, serious practical challenges often limit or preclude sponsors' ability to conduct target patient studies for new

and me too markers. Studies in the target patient population require access to well-characterized specimens with accompanying clinical data, while characterized serum banks are rarely available.

In the absence of bank specimens, prospective studies are the only source for the needed patient samples. The natural history of cancer puts these studies beyond the bounds of sustainable time and cost for most IVDs, especially me too tests. These products cannot support multimillion dollar studies lasting three to seven years.

In conclusion, on behalf of the HIMA member companies that manufacture these products, we ask this panel to recommend to FDA that the published guidance be followed; that substantial equivalence via comprehensive correlation studies described in Section 6A of the guidance are sufficient to clearly establish substantial equivalence of me too markers, and another legally marketed test for the same analyte in the same matrix with the same intended use and indications for use as another legally marketed test.

We recommend that FDA not persist in requiring target population studies for me too markers, because these studies do not make me too tests better, more useful, only more scarce and more expensive. Thank you for the

opportunity to address the panel.

DR. LADOULIS: Thank you. There is opportunity for a few questions now or at the conclusion of the comments. Is there any member of the panel who would like -- Dr. Jordan.

DR. JORDAN: I just want clarification -- this is Dr. Wilbert Jordan -- on what is meant by a me too marker. Are we talking about one that is exactly the same, measuring the same or something that is similar but not necessarily measuring the exact same marker with the same analyte?

DR. LADOULIS: Well, I think that the presenter probably ought best to define --

DR. JORDAN: That is why I wanted to present it to her to respond to.

DR. LADOULIS: What you mean, Carolyn, by -- and what you characterized as me too markers.

MS. JONES: A me too marker is measuring the same analyte in the same matrix as one that has already been approved. It is not -- for example, the list of markers that Peter put up in his presentation, if a manufacturer were to develop another -- a marker to suit one of those, it would be considered a me too marker.

DR. LADOULIS: Are you referring to the same

analyte such as prostate-specific antigen, as an example?

MS. JONES: Yes.

DR. LADOULIS: Okay. Any other questions at this point? I think then we might come back. If there are further questions she would be glad to answer them, I am sure. The next presentation is scheduled by Chris Zalesky of Centocor, Incorporation.

Agenda Item: Centocor, Inc. Presentation

MR. ZALESKY: I do not usually have too much of a problem being heard, but if anyone has a difficulty, let me know, and I will try to speak up. I am Chris Zalesky. I am responsible for regulatory affairs and I work for Centocor and to a degree we are the cause of today's meeting, I guess, by having filed a reclassification petition a few years ago.

I am pleased to appear before you today and would like to thank FDA for the opportunity to contribute to today's discussion, and I would also like to thank the members of this panel for continuing to volunteer their time to review this matter and other important matters throughout the year.

It is fitting that today's meeting coincides with the one-year anniversary of the reclassification of tumor

markers used for monitoring purposes. As the sponsor of the reclassification petition, and one of the first companies to have a marker undergo 510(k) review, we can say that we were very pleased with the overall results that we have seen to date.

As anticipated, reclassification has permitted both FDA and industry to make better use of limited resources by permitting substantial reductions in the administrative burden of the PMA process. At the same time, if these burdens are reduced, FDA has maintained the ability to apply a flexible and appropriate level of pre-market regulation to these products.

You also may recall that prior to reclassification some expressed concern that this change would unleash a flood of new tumor marker tests, and these tests would not be subject to an adequate level of scientific rigor.

So far, the flood has failed to materialize and certainly the products which have been cleared have undergone very rigorous review, indeed. In fact, FDA's efforts to assure scientific rigor have led it to depart from the guidance, which was the product of a collaborative effort between the Agency, this panel, the clinical laboratory and medical professions and industry.

The FDA has asked sponsors of second, third, and nth of a kind of so-called me too tests, to conduct studies in the target patient population. These studies are not required under the guidance document, and it would cause some difficulty for FDA and sponsors alike.

As a consequence of the request for target patient studies, FDA is faced with somewhat of a backlog of submissions and many sponsors have not been prepared or able to comply with FDA's requests, because the market for these products simply cannot support the cost and time needed to perform these studies.

In our experience even when sponsors are able to conduct target patient studies, the studies have not provided the Agency with any further insight into assay performance, relative to its need to make a substantial equivalence determination and they are not likely to provide the intended users with useful information upon which to assess or compare assay performance.

We are not here today to engage in a parochial debate of the merits and limitations of these studies. More important, we are here to offer our perspective that reclassification has so far been successful, and that the current guidance document describes reasonable and

appropriate studies and to support FDA's expressed interest in further improving the process.

To that end, I will touch on the following issues. First, what constitutes a new marker? Second, why is the current guidance adequate? Third, what key limitations of target patient studies are there for me too markers? And finally, what are some of the alternatives that we should consider?

Regarding new markers. When viewed objectively, a marker is only new once. When a sponsor is unable to identify another legally marketed test which measures the same analyte and the same matrix and has the same intended use and indications for use, this test as a new marker. Thus, as outlined in the guidance document, when the sponsor claims substantial equivalence to another legally marketed test, the test is not new. This is true, irrespective if the marker is a second, third, or 20th of a kind reviewed by FDA.

With respect to the adequacy of the current guidance document, as I noted, the tumor marker guidance was included in our petition as the central element supporting reclassification. Since then, it has undergone a number of modifications and is now the product of a collaborative

effort of FDA, this panel, the clinical laboratory, and medical professions, and our colleagues in industry.

Beyond serving as a designated special control for the purpose of reclassification, the studies described in the guidance were intended to relieve both FDA and industry from one of the primary inefficiencies of the PMA process for tumor markers. That is, the need to design, conduct, and review clinical studies in the target patient population for me too tests.

At best, these studies simply serve to reprove that which has already been established by the first of a kind marker. At worst, they contribute no useful understanding or new information regarding the performance of a me too marker, yet consume significant industry and FDA resources. While we respect FDA's interest in assuring scientific rigor, we were nevertheless surprised at its request for target patient studies.

It was our understanding that these studies described in the guidance document were reflective of FDA's expectations for tumor markers and other Class II IVDs. It was based on FDA guidances for other products as well as the accumulated history of 20 years of tumor market PMA submissions.

The studies described are adequate to assure the comparative safety and effectiveness of me too markers. The guidance describes target patient studies only from first of a kind markers because such studies are necessary to evaluate the performance of the test for the new analyte, and a new matrix, or when sponsors claim a new intended use or new indications for use.

Beyond supporting the claims of substantial equivalence, the guidance describes the studies needed to support the performance and other claims included in the labeling of tumor marker submissions, including the expected values, the reference range or cut-off, analytical performance of the test by comparison to another legally marketed device. Precision, recovery, linearity, and assessment of interfering substances.

As with other Class II IVDs, these data provide both FDA and users with significant useful information upon which to assess substantial equivalence and expected assay performance.

Moving on then to the key limitations of target patient studies for tumor markers. Tumor marker studies have both practical and scientific limitations. From a practical perspective, studies in the target patient

population require access to well-characterized specimens, with accompanying clinical data.

The FDA has always made it clear that it is willing to accept studies from well-characterized serum banks, however such samples are rarely available and when available they are very expensive. In the absence of bank specimen, prospective studies are the only source for the needed samples and data.

The natural history of cancer puts these studies beyond the bounds of sustainable cost and time for most IVDs, especially me too tests. These products simply will not support multimillion dollar studies lasting three to seven years.

From a scientific perspective, the nature of clinical studies of tumor markers generally limits the value and validity of study-to-study comparisons of clinical performance. Indeed, comparing the results of two independent studies conducted in a target patient population using two different tumor marker products may yield results which are similar or different.

When the results are similar, this does not necessarily support the conclusion that the two tests have equivalent clinical performance. Conversely, when the

results are different, this does not necessarily support the conclusion that the two tests have different clinical performance.

For tumor markers, the similarities or differences between tests in terms of sensitivity and/or specificity are only meaningful if they are observed when evaluating the assays -- both assays -- in exactly the same study population. In our experience, this is because there are a number of uncontrolled or uncontrollable factors which can have a significant impact when studying tumor marker performance between studies and the target patient population.

These include, first of all, availability of first of a kind markers impact the ability to perform studies of me too tests in the same target patient population. For example, use of CA 125 has largely replaced second look surgery in monitoring ovarian cancer patients for residual disease following primary surgery. It is therefore not possible to conduct an independent study comparing the results from a me too marker to the clinical findings from second look surgery and this target patient population.

If FDA deems these studies to be a clear requirement, then Centocor CA 125 2R1A will remain the sole

test available in the United States with a second look, intended use client.

Next, tumor market studies are impacted by the availability of alternative approved or unapproved IUO or laboratory home brew marker assays. This influences the characteristics and make up of a study population and confounds study-to-study comparisons. Excluding access to these tasks in the study protocol certainly improves control but creates a selection bias, because only patients who would not likely benefit from the available markers would be enrolled in the study.

Also, tumor marker studies are impacted by other factors which limit the value and validity of study to study comparisons. First among these is small study size. Unlike most other clinical studies involving IVDs, enrollment in prospective multi-year, multi-site tumor marker studies, directed by well-established clinical researchers will net only small numbers of valuable patients, comparing the performance of the same or two different markers in two small studies. Quite large apparent differences can often be observed as a consequence of small study size.

Uncontrolled differences affecting even one subject can have a disproportionately large apparent impact

on the results observed between two studies. Also, study protocols are unavoidably impacted by treatment and therapy decisions. As you all know, tumor markers must necessarily be studied in patients undergoing treatment, or be monitored for detection of recurrent residual disease.

Treatment and therapy and monitoring decisions, including sample collection intervals and changes in treatment protocols are usually dictated by the needs of the patient, without regard to the impact that variations of these factors may have on the results of the tumor marker study.

These factors result in uncontrolled and potentially significant differences between study populations. Finally, studies often extend beyond the time frame for changes in medical practice. Multi-center studies intended to follow sufficient numbers of patients to a predetermined clinical end point can last two to five years or longer.

These prolonged time frames may match or lag behind advances in treatment and therapy, and prolonged study time frames contribute to uncontrolled population differences over time, thus impacting the validity of study to study comparisons.

All of the foregoing factors represent uncontrollable differences which limit the validity and use of study to study comparisons for tumor markers. Only side by side comparisons in the target patient population will control for these fact errors, however such studies provide little, if any, useful data beyond that already provided by the side by side studies described in the guidance document.

Considering our view that the current guidance document is adequate, and the target patient studies have very limited scientific and regulatory value, we would like you to consider a few alternatives.

First, we can do nothing. However, this would not solve any of the current difficulties facing FDA and sponsors.

Second, we could simply comply with the current guidance document as it has been written. This would be acceptable to Centocor and likely to most industry sponsors, but it may not address FDA concerns regarding second, third, and fourth of a kind markers.

Third, we could consider revising the guidance document to minimize the current difficulties, or to take into account other sources of information; for example, can the limitations of target patient studies for me too markers

be overcome, and if so, how?

Based on our own experience, I am not sure that it is worthwhile to continue to invest further time and effort in target patient studies, nor to revise the guidance to include these studies as a requirement. However, perhaps a panel of well-characterized serum samples would be an appropriate alternative to target patient studies.

While this is a considerable undertaking, such a panel would serve to provide additional structure for a comparative evaluation. Once developed, that panel could also be used as a form of manufacturers' proficiency testing program, and perhaps minimize or diminish assay to assay differences in standardization.

While it is not clear to us that additional scientific rigor is needed to establish the substantial equivalence of markers, what role, if any, can post-market studies interpreting clinical practice guidelines, or the existing body of literature, play in this evaluation?

As we heard this morning from Dr. Gutman, in its ongoing re-envision efforts, FDA has proposed some very creative approaches to approving the efficiency of the pre-market review process, which benefit the Agency and industry alike. These proposals often diminish the reliance and

submission of device-specific testing and evaluations in favor of recognition of established standards or other forms of data.

In view of the limitations of the target patient studies, and the success so far experienced with its re-invention efforts, would it be appropriate to evaluate the role of post-market studies, clinical practice guidelines, or the substantial body of literature supporting the use of me too tests for monitoring?

In summary, we would like to suggest that you consider the points that I identified above there, and we would certainly be pleased if you agreed with our perspective that the current guidance document is adequate. In our view, it describes appropriate studies for both new and me too markers. For me too markers we see no scientific or regulatory basis for requiring additional independent evaluations in the target patient population.

In our experience, such studies do not add any further insight into assay performance relative to making a substantial equivalence determination, nor do they provide the user with information upon which to assess similarities or differences between two assays of like type.

If there is a need for increased scientific rigor

beyond that already provided in the guidance, it may be worthwhile to explore the use of well-characterized serum pools, or to rely more heavily on the clinical practice guidelines now under development or in post-market studies, as well as the significant body of information available in literature. We appreciate the opportunity to share our thoughts with you today. Thank you.

DR. LADOULIS: Thank you, Mr. Zalesky. Are there any questions from members of the panel?

MR. ZALESKY: Thank you.

DR. LADOULIS: One question. Do you have -- do you or the other trade representatives, such as Health Industry Manufacturers Association, undertaken any serious proposals for investing jointly with other manufacturers in the development of such serum banks?

MR. ZALESKY: We have in sort of a modest sense. Centocor itself actually has a program that is available to companies that buy our antibodies and antigens, and for lack of a better description, at this point we are calling it a harmonization program, not a standardization program.

In essence, it sort of serves as a manufacturers' proficiency testing program, but it occurred to me in the course of us putting that together -- and I have spoken to

Peter and I have spoken to the folks at AACC just on a preliminary basis -- that sort of evaluation might be something that could provide us with information we certainly do not already have, and serve that comparative purpose that we all want and need without the need to do target patient studies, and at the same time too, might help to diminish differences in standardization among the available assays.

DR. LADOULIS: Well, maybe we will hear more about that to follow as we have our next invited presentation from American Association of Chemical Chemistry.

DR. CHAN: Right, thank you.

DR. LADOULIS: Dr. Chan. Thanks very much, Mr. Zalesky.

Agenda Item: American Association of Clinical Chemistry Presentation

DR. CHAN: Dr. Ladoulis, Dr. Maxim, Dr. Gutman, members of the Immunology Panel. First, I would like to thank you for the opportunity for me to come speak to you this afternoon. My name is Daniel Chan, I am Director of Clinical Chemistry at Johns Hopkins Hospital. I am also Associate Professor of Pathology, Oncology, and Urology at Johns Hopkins University School of Medicine.

I am here to testify for the American Association for Clinical Chemistry. I personally do not have any financial involvement or interest in any of the products or any companies, however my laboratory at Johns Hopkins has received and we currently receive many research support from all the companies that are involved in tumor markers. I am currently conducting research on most of the tumor markers that we are interested in.

The AACC is a professional society that consists of clinical chemists representing approximately 11,000 professional laboratory scientists. We work in the hospital in independent laboratories and in diagnostic industry, nationwide. As laboratory scientists, we lay an active role in tumor marker research, development, and refinements of many laboratory tests.

We also use serum tumor markers for diagnostic, for prognostic, and for monitoring purposes, and as an association we strongly support the FDA's reclassification of serum tumor markers from Class III to Class II devices.

Given our members' extensive involvement in tumor marker testing, AACC would like to offer the following comments on the FDA's guidance document for the submissions of tumor-associated antigen pre-market notification, namely,

the 510(k)s.

First, I would like to state that AACC fully supports the Agency's September 19, 1996 guidance document. We believe that this document is well-written, concise, and provides appropriate guidance to the manufacturer, so that they can easily understand what information is required for the FDA for such devices.

We are pleased to note that as a result of this document, several additional tumor markers have been approved in the last year, as stated by Dr. Maxim this morning.

As a professional laboratory society, we believe in full disclosure of analyte contents and testing methodologies. For example, detailed information should be provided on the nature of the antigen, the antibodies, and their bindings to the specific defined epitopes on those antigens. Using defined patient populations, and with appropriate statistical analysis of the data, test interpretation should be provided to evaluate clinical outcomes.

This data are useful, not only for seeking FDA approval, but also in the marketing of the product by the manufacturer. Clinical chemists like us routinely seek this

type of information when we evaluate a particular tumor marker assay, in trying to decide which part that we should use for our laboratory. We believe that this document provides appropriate guidance for such an FDA submission.

We would also like to briefly comment on the issue of me too tumor markers. AACC agreed that such tumor markers should be subject to less stringent submission requirements, however, we encourage manufacturers to provide sufficient scientific and clinical information to prove that, indeed, those markers are substantially equivalent to the first approved marker.

We realize that this is not always an easy task; for example, from time to time, it may be necessary to demonstrate that common epitopes on the antigen are being measured by the different assays. Providing the FDA with sufficient clinical data will only speed up the approval process, as well as the introduction of the new markers to the marketplace.

In summary, AACC strongly supports the FDA guidance document, and the reclassification of tumor markers. We believe that good scientific and clinical studies, as outlined in the guidance document, will lead to greater understanding of the use of tumor markers and

ultimately to improved patient care.

I thank you for the opportunity to speak before the panel today, and if you have any questions, I will be happy to respond to them. Thank you.

DR. LADOULIS: Dr. Chan?

DR. CHAN: Yes.

DR. LADOULIS: Could you elaborate on one of your statements and that is, the need for demonstration of information about the characterization of antigen and antibody specificity. Is it your explicit or implicit statement that such information is not adequate now?

DR. CHAN: I am not implying that such statements are not adequate. I am saying that from the scientific point of view, trying to determine if tumor markers are equivalent, really cannot just be based on the antigen itself; you need to look at the antigen, the antibody that is measuring the particular antigen, and the particular methodology. All that information will affect the test results and therefore the clinical outcome. And so that is what I meant.

DR. LADOULIS: You mean on product description information that is available to the laboratories, is that what you mean? Or do you mean to the FDA agency that is

evaluating --

DR. CHAN: I am suggesting that that information should be provided to both FDA and the clinical laboratory for us to evaluate the performance of those assays.

DR. LADOULIS: The users of the --

DR. CHAN: The users, as well. I am saying that the same information could be applied, or could be given, to the users. So, basically, I felt that the manufacturer could do one study, and basically meet all the requirements for either regulation or marketing purposes.

DR. LADOULIS: Okay, so that the question I have for you is that, from the user's perspective, representing the clinical chemists and laboratorians, do you feel that the information that has been provided up until now for products that are in the market, have sufficient information for you to evaluate the substantial equivalence from your perspective?

DR. CHAN: I was saying, from my perspective, in some situations, they are; in some situations, they are not. I would say in most situations, they are. I do not know whether I should go in to further give you examples. If you would like to, I will be happy to.

DR. LADOULIS: Unless there is some particular one

glaring example that you would like to present, one particular antigen, without this being specific for any particular manufacturer or product, if you think there is one category of tumor marker that deserves particular scrutiny.

DR. CHAN: Well, speaking for myself, not for the Association, I could give you an example of PSA as an example. As we now know that PSA is more complicated than before, and the different antibodies may recognize different forms of PSA differently, so that unless you clearly define what antibody, what epitope you are measuring, you cannot really evaluate the outcome, and that is one of the examples that I was trying to explain.

DR. LADOULIS: Thank you.

DR. REYNOSO: I have a question.

DR. LADOULIS: Dr. Reynoso.

DR. REYNOSO: Professor Chan, would you care to comment -- do you have, or are you in a position to state your position or the AACC's on the issue of additional clinical studies on the target populations? You did not speak to that point explicitly.

DR. CHAN: Additional clinical studies, additional to --

DR. REYNOSO: On the target population.

DR. CHAN: In addition to what?

DR. REYNOSO: Clinical studies in the targeted population.

DR. CHAN: I guess I do not quite understand your question.

DR. LADOULIS: Previous speakers have addressed the concern that there is a requirement for targeted population studies, and that this may be a burdensome requirement, and may not actually be a requirement that is consistent with the guidance document.

You have, therefore -- in response to this question, that is the issue, I think. It is the additional -- for additional submissions for products that are claiming to be substantially equivalent. Do you think that clinical studies in targeted populations are still advisable or called for?

DR. CHAN: Well, as I was trying to address the issue of me too markers, and my feeling is that the manufacturer has the burden to prove that it is indeed substantially equivalent. One of the ways to prove that they are equivalent is use certain target populations of study to do that.

I am not suggesting that you have to do prospective clinical trials. Extensive studies like was mentioned by the speaker, may take two to five years, those kind of studies. But, I do think certain clinical information is needed to define whether they are equivalent or not. Simply to do an analytical correlation in my personal viewpoint is not sufficient to show that they are the same. I do not know whether that answers your question.

DR. REYNOSO: Yes, partially it does. Thank you. Perhaps later you can elaborate on what you have in mind when you say, perhaps others, rather than a full clinical study. Maybe we can talk about that later.

DR. CHAN: Sure. Yes.

DR. REYNOSO: But I understand your answer, thank you.

DR. CHAN: Thank you.

DR. LADOULIS: Are any other panel members who have any other questions of this presenter or others?

DR. JORDAN: No. I have a question. I am wondering why there is no one from any of the medical associations, particularly oncology, addressing this issue.

DR. LADOULIS: Is there anyone from -- representing an oncology association in the public attending

that would like to make any comment? Appearing to be none. Any other questions? Well, I think that concludes the open public hearing, then.

If there are no further questions or comment from the public, turn this back to comments from the Executive Secretary.

Agenda Item: Question One

DR. MAXIM: Okay, thank you. As I mentioned this morning, what we are going to do is present slides, questions or outlines of questions that we would like some basic information on, or the panel's input on these different questions and see if you can offer us any advice as to how, without changing or rewriting the guidance document, which we are not proposing to do, but at least define a little bit better where we want to go with it.

This question is in your information pack, also. And essentially it says, what do you recommend as data requirements, or what would your suggestions be for data requirements for Class II tumor markers, considering and arbitrarily breaking the markers into the following scenarios?

You have tumor markers such as CEA, AFP, and we put PSA into that group, also. We have a significant Agency

history with them, having evaluated in the neighborhood of 10 to 15 different submissions. There obviously is quite an extensive literature base on these products, as well as clinical history and clinical experience with them.

We compare that type of marker, or that family of markers, with tumor markers such as CA 27.29, and CA 15-3, where we may only have one FDA-approved product for each of those, and I believe it is more than one now, but we have minimal Agency history or experience with it. There is a variable literature base, not having had that much experience with some of these markers. Clinical history, the laboratory experience is not all that extensive, in some of these cases.

The other one that we mentioned as a category, as a scenario, would be a tumor marker such as CA 125, with -- Mr. Zalesky mentioned this morning, we have approved that one time back in the mid-eighties for the second look surgery indication.

We have really seen no data at the FDA with regards to this monitoring recurrence or response to therapy from any manufacturer, and although there is a considerable literature base, you can certainly find a use for -- various uses for the marker in the literature, and a fairly

extensive clinical history where it is recommended and I believe it has come up as a prior standard of care, practice of medicine in some areas.

Basically, with Dr. Ladoulis' help, I would like to go around the table to each of the panel members -- could you just leave that up for the record, please -- and get some of your comments on these various groups of markers, these various types of scenarios.

What we are looking at -- and again, to clarify with this targeted patient group -- we are essentially saying that if a manufacturer were to market a product that is indicated for monitoring Stage II and III cancer patients for recurrence, as the middle scenario, do you feel that, in addition to a side by side comparison with a legally marketed predicate device, that perhaps reactivity of the device with normal patients, breast cancer patients, other cancer patients, benign diseases --

Do you feel that there should be a responsibility put on the manufacturer to come to the Agency with data to support that as a medical claim? That this product has a certain sensitivity and specificity and certain clinical performance in its ability to monitor and look at Stage II and III breast cancer patients for recurrence, for the

proposed intended use. That is the targeted patient population that we have been referring to the biggest part of the day.

DR. LADOULIS: Does everybody understand the target population definition? Yes.

DR. MCCASKILL-STEVENS: I guess this is a question and a comment. I think that, in my mind, if you are going to answer the questions about subsequent tumor markers, me too tumor markers, a lot of burden lies on the primary approval. The target population at the initial approval is going to be very important.

One of the things that I had noted in the past, was there have been some issues with actually the population, whether it has been inclusive enough to represent a large enough population, be it special population or even if you are talking about sample size.

I think if we are going to address the question of subsequent approval, there is a lot of burden on what is going to be done up-front. If you do not have adequate representation or adequate numbers when the approval is given, then it is going to be very difficult to give me too approval down the line.

One other comment I would like to make is, in

reference to the interactions between treatments and the previous treatments, one argument that one might make as an oncologist, is that there are interactions with drugs, and it may be that one might question whether one would not want to do another clinical trial because of the various environments and armamentariums in which one has treatments.

I mean, we have made some progressions and certainly evolutions in terms of our treatments since the approval of some of the tumor markers that are even listed on this chart. I mean, I would just like to start out with that comment.

DR. LADOULIS: Thank you. Any other comments? I think that one of the qualifications to your statement is important, and that is that you pointed out, for the monitoring of established diagnoses of cancer. Because as you look at these markers, the one that has stood out as the exception in its intended use has been prostate-specific antigen, or PSA.

It has been approved for the purposes of diagnosis, in addition to digital rectal examination. That makes it a singular tumor marker compared to all others. And all of the others have been approved for one purpose only, and that is, for the monitoring of already diagnosed

cancer.

I wonder if this concerns the members of the panel as to me too products that might be for monitoring, versus those that might be for intended claim of diagnosis.

DR. TAUBE: Sheila Taube. But when we say, me too, are we including the indication? Because if you say it is for exactly the same indication as what a previously-approved device indicates, and then you give all of the information about the performance, that is a different issue from saying -- I mean, when we looked at PSA again, it was not for the same indication as the original approval. So, if the indication is the same, I think that it is different if it is identical.

I also think -- I mean, one of the things that Worta brought up about the interactions with drugs. Even if you do not do an entire clinical study, since the drug regimens change, I think that at least there has to be some information, there have to be some samples from patients receiving some of the new drugs to indicate that the test is still valid in that setting. I mean, even for monitoring, supposing it changes the performance of a particular device, or marker?

DR. LADOULIS: That would be for both the

previously-approved as well as the proposed --

DR. TAUBE: Yes. Yes. I mean, so the side by side study would be useful.

DR. REYNOSO: Yes, I would like to speak to that subject, because I have suggested something similar in another context and that is, if you follow a previously-approved marker over time, and follow the performance, and there will be ups and downs and correlations with responsive therapy and changes in therapy and so forth.

One of the meaningful things we could ask is to do at least some studies showing that the me too test follows the same pattern. That would be very good evidence of the special areas of equivalency. Because it would take into account some of the changes that Worta was saying. In other words, following over time, and I have suggested this in another context, do I make sense?

DR. LADOULIS: In the same individual?

DR. REYNOSO: In the same individual, correct. Yes.

DR. LADOULIS: I think that is a good point.

DR. AMMIRATI: Doesn't that bring us back to the target population, however is that what you are going for, to see how someone responds to therapy over time?

DR. REYNOSO: Yes. May I answer that? Yes, I would say, in a more limited than a full clinical study, but rather follow a number of patients. I am not prepared to say how many, but a number of patients over time, to show parallely, as the patients change and therapy changes and drug interactions occur, to show parallelism of response over time, between the established approved marker and the proposed new one. That is another way of showing clinical parallelism, if you will.

DR. AMMIRATI: Right. I think the problem we found in industry is that if you -- well, you have one of two choices. You can either bank samples, which have rather well-characterized histories and we have seen the clinical course of the disease, and we can track up and down, and they are hard to come by and very costly.

The problem we run into in a prospective study is that you may not get the same results, because you do not have the same person, in the same course of time. Getting back to some of the interference testing and clearly with the therapeutics changing over time, I think a lot of these studies could be done in the laboratory, in vitro, and NCCLS has a very good document under EP-7, I believe, talks about interference testing, and there are data that can be

provided where you -- in vitro, in an in vitro manner, introduce metabolites or the drug itself into various matrices and we can generate data that way, and I think that ought to be considered. Because that is a good point, that treatments will change and there could be -- a current test would not be aware of a possible drug interference.

DR. LADOULIS: Go ahead.

DR. HORTIN: I am a little bit disturbed about the break-up of these into the three categories of basically how familiar you are with the marker. It seems to me that that may not be, from a clinical standpoint, the most useful break-down in that it seems to me if we are looking at kind of risk to the patients, the primary factor is going to be, what the clinical response is going to be to the marker result, and if you have a clinical response, say to maybe a CA 125, but you are going to do an abdominal ultrasound, that is not necessarily a very high clinical risk for the patient. It is going to require a non-invasive procedure.

If your response to results is going to be a biopsy, or a second surgery, or a relatively invasive high risk procedure to the patient, it would seem to me that that is where the clinical risk is involved, so I have not ever seen -- it seems to me more appropriate in some respects to

break these down in terms of the risk and the clinical requirements based on what the clinical impact and kind of the clinical response is going to be, and I have not -- whether it is a marker that you have a lot of experience with, but it is a high risk marker.

It would seem to me that there would be a more clinical burden for more clinical information if it is going to be a low risk intervention. Basically, it would seem to me that the risk is not going to be so great, and it may be useful to stratify in that fashion and to some degree the classification, whether it is Class I, II or III is based on those considerations, but it seems to me that should come into play somewhere.

DR. MAXIM: I think we would agree on that to a point. Obviously, we are concerned with the risk of the markers, how they are used in specific applications. More falls to the use of the markers, the practice of medicine afterwards, and basically our reviews have to be focused on whether or not the device does what it says it does in the intended use population, so we pretty much have to draw back to that.

If there was a high risk element associated with the use of it I think is something we consider. This is a

very arbitrary break-down of the different types of markers, and FDA experience with them was just one parameter we could use. I am sure there is probably 100 other ones that we could have incorporated, but I chose this one just to bring it up for discussion today.

DR. LADOULIS: Dr. Kemeny.

DR. KEMENY: Another problem with target population is, what you exactly mean by that, because I remember us going over some markers and we were actually speaking about this at lunch, where there was not, for instance, a good ethnic diversity in the target population. Are we going to be talking about that, or are we just going to be talking about, like, breast cancer patients.

Maybe that needs to be defined a little bit more clearly, and also, the fact is that we do not have a lot of data on the markers that we have in reference to some of these questions. We were going to get some post-market information that maybe we did not always get on some of these things.

DR. MAXIM: In the one case, the post-market study that you recommended is still ongoing, and the company is to provide that information. I think what you need to focus on, also, is the fact that the first time, a lot of the

questions -- Dr. McCaskill-Stevens brought some of these up, as well that you just alluded to.

For the first indication, or for the first use of this in a particular population, the panel will be involved and there definitely will be clinical information, a clinical study associated with the clearance of that marker for that particular intended use.

You would have an opportunity at that time to evaluate all of these parameters. Whether or not the study size was sufficient. Whether or not the samples were sufficient. Whether or not the demographics and the patients were represented appropriate. Whether there were demographic or ethnic variations that were not taken into consideration. The effect of interfering substances. Whether or not the patients themselves represented the current method of how the patients are being evaluated at that point in time. And even make recommendations for limitations. As new chemotherapeutic agents were to come to the marketplace, these things should be reevaluated.

What we are looking at is what happens after that, having established that, having established the optimum characteristics for the use in this, and having characterized the target population for its use. What does

the Agency do the next time? Is the second one then -- the second marker does not have to fall back and even look at the patients at all?

I mean, is it sufficient that they just come in and not even evaluate recurrence in Stage II and III breast cancer patients at all, but they can take that intended use? Should that happen? Should we start looking away from that on the fourth one? On the tenth one? Should we always have some sort of a correlation that, if this product, if this manufacturers says they can do that, that is their medical claim, their indication for use, that we in fact should see data, see clinical data in their submission to support that.

DR. HOMBURGER: May I comment on that?

DR. LADOULIS: Sure.

DR. HOMBURGER: This is Dr. Homburger. I am a little reluctant to comment at all. This is my first meeting and I am not familiar with what is in the guidance document, but I suppose unless you want me to sit here for four years and say nothing, I should break the ice.

I actually think that in part the answer was given, I believe, by the speakers who presented. I think Dr. Chan actually implied a very logical way to go about this. To the extent that a particular marker as a

biological analyte is very well-characterized; in other words, the target antigen is well-characterized, the number of epitopes are defined, and the immuno-chemical detection reagents, in this case almost invariably antibodies, are well-characterized and defined.

More often than not, they are monoclonal, more often than not, one could precisely define the antigenic site, to the extent that somebody else comes along and says, we are going to measure the same thing, and we are detecting the exact same antigenic site, and here are the characteristics of our antibody, here is its affinity, etcetera.

I think you can almost not do clinical studies at all, unless you fundamentally change the immuno-chemical approach to measurement. Because you are measuring -- you know, by definition, that you are pretty much measuring the same thing and given the degree of variability that you are likely to see in clinical studies, from patient populations and the other factors that various people have mentioned, you are not going to learn much more from doing that. However, take another case. A highly heterogeneous marker with multiple antigenic sites, multiple different antibodies that have been used and described in the literature to

measure that. You do not know necessarily all that much about the metabolic fate of this antigen that you are trying to measure.

In that circumstance, I do not think you can simply conclude that looking at two different epitopes with two different antibodies and a complex molecule, that you are going to come up with two markers that are going to behave in the same way clinically.

If I were to offer a suggestion, I would recommend some stratification of the demonstration of equivalency based upon what we know about the immunochemistry and the metabolic behavior of the marker itself. And I think that information is something that maybe should be provided, when somebody says we are going to measure analyte X, or tumor marker X.

DR. MAXIM: Thank you.

DR. LADOULIS: I think the history of the assays for CEA, for example, illustrates exactly what Dr. Homburger pointed out, in that the evolution of the clarification of the number of different epitopes and the heterogeneity and the complexity of a marker such as CEA, which was not originally well-characterized and became more well-characterized with time, illustrated that that was the basis

for the variability between different vendors' products in terms of the assessment and performance which came out in the literature later.

I think to a certain extent other antigens may be less complex than surface glycoproteins, and may be more easily characterized, but I would agree with what Dr. Homburger just pointed out, that to the extent the antigenic marker exists in the matrix, in the human population, in a form that is fairly well-characterized, has a finite number of epitopes, and specificity of the reagent that is being used has been identified for a particular epitope, and is substantially equivalent to another combination of antigen and antibody by a different previously-approved device. That would constitute, I think in my mind, substantial equivalence, but that may be a minority of instances. I do not know.

I think the other qualification might be just what Dr. Reynoso suggested, that a certain select small series of measurements in the same individual patient by a already-approved device and a candidate device for approval, showing the same trends, maybe where the systemic bias having to do with different affinity constants for the antigen antibody reactions for the same epitope, that might constitute

substantial equivalence, too.

One of the questions that I have is whether or not there are several major issues to be identified should, number one, the target population -- and if anybody wants to make any comments about that first, I mean if -- I think that would be one of the foremost questions that is being faced here in view of the comments from the public and the industry.

Second, does the FDA have in fact any inconsistency internally between the practice and the document, the guidance document, as was suggested in one of the first comments?

Third, is there a need for having a bank of serum markers for certain of these tumor markers?

Finally, is full disclosure of the antigenic and antibody characterizations adequate now, as it is required for compliance by devices that are approved and is it being required sufficiently in those that are in the process? Those are the questions, and anybody want to make any comments about those?

DR. HOMBURGER: I would just like to express one concern and that is that, for a given tumor marker, that is approved for use in a relatively restricted clinical

setting, it strikes me that the clinical application of those measurements is going to evolve over time in response to what is going on in clinical medicine. If there are better treatments to offer, or alternative treatments to offer at some point in time, that marker may become extremely useful in a setting for which the original application had no -- where there was no consideration of that.

What do we do under those circumstances, because for sure, physicians are going to read the literature, they are going to use it for that, it is going to be marketed for that. Does FDA at that point in time, since they are not necessarily aware that this is going on, I mean, do they just put the blinders on, or how does this whole process work?

DR. MAXIM: That happens fairly frequently across the FDA, I am not just speaking for tumor markers alone. We are well-aware of the fact that even with drugs, as drugs are approved and put out, many of the physicians can come up with innovative uses of the drug itself or combinations of the drug. Fen-fen is a classic example, the most recent example that comes to mind.

The same thing happens with the tumor markers.

The Agency does not pretend that the PSA assays that are being cleared for use in monitoring prostate cancer patients are being used strictly and solely for that use, but obviously that they are used off-label in the laboratory for the early detection claims or for screening or whatever you want to call it.

The manufacturer, our regulations and our rules call for the manufacturer should not be promoting for that particular purpose, and that they should only promote within the scope of their own intended use. But, no, there is no way that the Agency has found in the past to be able to deal with off-label use or creative uses of previously-approved drugs. Not so much biologics, I think, but certainly devices.

DR. HORTIN: If I might then follow-up, I would just suggest that -- make another pitch for more rigorous immunochemistry in the description of these products, because that provides a sound underpinning to whatever clinical uses ultimately get approved.

If we know that physicians and the population are going to demand access to this in a particular medical setting -- and we want to make sure, or provide some assurance, that those products that are out there are

substantially equivalent -- I do not think we can go back and ask to do a clinical study every time somebody extends the use of something.

The only real assurance that somebody would then have would be substantial analytical equivalence, and I do not know that enough information is being provided right now to assure substantial analytical equivalence, but if it is not, I think that might be something that deserves further consideration.

DR. TAUBE: I just wanted to comment and in particular with the tumor markers in breast cancer, we do not want to continue to approve me too markers that have lost their utility, and that can happen as well.

I mean, we were testing in a particular adjuvant setting, and if our drugs are evolving and transforming from our metastatic setting into our adjuvant setting, then that, too may change, so I do not think we want to approve, you know, five, six and seven generation me too drugs that may have lost their utility.

DR. REYNOSO: I think that we ought to come to some clarification maybe at this point, in that in order at least for me to be able to continue monitoring the deliberation of this panel and to provide some of the advice

that the FDA needs, I think that we have to understand that Dr. Homburger made an excellent point in speaking about the need for a stratification, and that not all tests can be judged the same.

There is a substantial number of tests in which the approved method has been approved for a long time, the clinical data are available, and the me too tests are simple adaptations to a new instrument, or simplification of methodology or what have you, when in fact the antigenic source may be the same, and the antiviral source may be the same. And I think under those circumstances, I can see why analytical equivalence should be enough, and one could require additional clinical studies under the setting I am describing.

There may be tests when that is not the case, so I think it is important to come to grips with the concept that we cannot answer the question being forced by the FDA and by the sponsors, in terms of all of the tests, but some degree of a stratification, so that we agree that under certain circumstances, analytic equivalency may be sufficient, and in other circumstances, additional clinical data may be needed, and I think that this point about the stratification is a very important one.

DR. MAXIM: Well, I agree, I think we could go with that, too, then if you would like to try stratify these tests along those bases. Keep in mind, also, that we can only look at the material that a sponsor would send to us, and you are going to have various sources of antigen, mixtures of antigens. Each company is going to have their own antibody or in some cases they share them.

I think we looked at data this morning to indicate that, if you are going from a test, and you are dealing with the same antigen and antibody, the likelihood of analytical comparison, analytical identity, between these two tests is better.

You cannot force manufacturers or force sponsors to adopt a single antibody using all their assays, and a single antigen, a single confirmation of the device. This is the uniqueness and the variability of each one of the products that comes to the marketplace. So you are going to drift along those lines even if there is analytical lines. We have dealt with this consistently since the beginning of approval or clearances of tumor markers by putting warnings in the labeling that each one of these devices does function differently, and does measure with various degrees of differences, and that care should be taken if you are

monitoring a patient or evaluating a patient before you change assays, that you should re-baseline that individual so that you can be comfortable with the two results getting from the new kit, the new device.

DR. LADOULIS: Yes, I would amplify that. I mean, for example, having the same antibody and same antigen and yet having different device characteristics such as a solid phase immunoassay, versus a liquid phase, has given differences in performance characteristics between some of those devices that have been approved for CEA, AFP, etcetera, and significant differences were suggested in some of those studies.

Are there any other comments about stratification or some other way to make some stream -- help the FDA to streamline this process in terms of evaluating the components for substantial equivalence? Is there anyone from the Agency that would like to make any comments?

DR. MAXIM: I have been commenting right along -- [simultaneous discussion] --

DR. LADOULIS: Is there anyone from who would like to comment about some of the statistical differences and the criteria?

DR. MAXIM: Dr. Rabinowitz asked, again, another

aspect which we have not addressed or dealt with is that you look at the industry point of view on this, also, what disadvantages then are you placing on the first manufacturer, the person who goes out and develops the assay, does the clinical study to support the usefulness of it, and puts it on the marketplace to have anybody else then simply come back in with the side by side comparison and the me too attitude, that they can just tag along and move along with minimum clinical data to support their substantial equivalence?

DR. JORDAN: I had to questions about that, earlier. One, where does patent play in with this? I assume one has to have a patent to go ahead. It is now so many years before they can do that, or does it matter?

DR. MAXIM: The patents do not really play all that much of a part in it. Obviously, each manufacturer would have -- if it was a patent in the form of the assay, they would have a patent protection for their product. But then the next person would come in and could move against that patent, or put their product on the market with their own format and perhaps a non-patented format.

DR. JORDAN: Looking at it from the role of the manufacturer, as a capitalist, I have put in money. I have

spent lots of money, I assume, with clinical studies, etcetera, to prove my product. And if for example, my competitor here has done nothing, and now comes along a year later and has not put anything out, when I have not even regained what I put into it. From the manufacturer's perspective, I think that certainly does seem unfair. Now, I am not a manufacturer, I do not know.

DR. MAXIM: Well, that was one of the benefits, if you would call it, of the PMA process, because you did the clinical studies. You would have PMA protection. You would have PMA approval for that particular device, and then as a Class III product, additional manufacturers would have to come in with similar types of data in order to go to the marketplace. So it gave you some leeway, some lead time in being able to market your product without much competition in the marketplace, other than that coming in from, say, the home brew arena.

I do not know that anybody would be in favor of taking a step back, simply for the protection of industry -- of the manufacturer -- to go back to where it would be a Class III situation, and we were not taking the benefits of reclassification. But again, this is the type of thing, when do you become me too, where the manufacturer has gone

through that and put that product on the market with a serious investment? Would it be sufficient for someone to come in, take that kit as it becomes marketed, run 100 serum samples and say, I am the next one?

DR. REYNOSO: May I comment on that? Yes, I think it depends on how the question is being asked. If the question is being asked in terms of scientific validity or clinical validity, or regulatory validity, then demonstrating analytical equivalency is perfectly fine.

If you are asking the question in capitalistic terms, is it fair to the other company? Well, I do not know that, but I do not think that is a concern of the FDA, necessarily.

DR. MAXIM: No, it is not. But it is a different -- it is not truly one of our concerns, but it is a different aspect that I would like to bring up for discussion.

DR. AMMIRATI: Yes, I think that -- I want to reiterate kind of where we start and stop in terms of -- certainly, everyone can have an opinion, but legally, some of these pharmo-economics(?) and some of these issues really go beyond the purview of what FDA can do and what this panel can do. And industry, just my own personal thought, really

understands that.

Being the first to the marketplace, although you do have to do more work in terms of one of a kind, or the first to market tumor markers, is considered -- you do not know how long you will be there. Usually it is more than a year. Again, that is independent of what we heard or discussed, but I certainly would not think that fairness to the industry should hold back the thought of supporting the guidances as is written. And if those people have the opportunity to go back and read, I believe 6A, what is required to bring even a second of a kind to market, is a -- it is not trivial. It is not, do nothing. It is still very comprehensive. It looks at reference ranges and certainly the performance characteristics you would expect to see in the labeling, and that is not trivial. But there is a lot of information that is made available.

DR. LADOULIS: I do not know if we lose focus, I would like to try to get back to focus on the questions about the guidance document, and the reclassification, if we could, and whether or not the comments that were first made this afternoon by the Health Industry Manufacturers Association that to me, as I understood it, suggested that there may be even some internal inconsistency between the

policy and the procedures, or this process.

I would like to hear some of that addressed by the Agency. If that is in fact a reasonable statement, or is not reasonable, and what response there might be? Because there was a suggestion there that there is some internal inconsistency in having this kind of a targeted population, and is that inconsistent in the way that the guidance document is actually formulated, and the way it is followed.

DR. MAXIM: I would start, and perhaps Dr. Gutman would like to comment after me, the guidance document was formulated as a -- not as a requirements for what it would take to get products clear by the 510(k) route, but truly guidance to the industry.

Our interpretation was that as we got started wit this process, we would go through the analytical characterization and various levels of looking at the performance of the device and the targeted population. Possibly any type of involvement at that level was more than the industry expected, which is the reason we are here one year later trying to figure out whether or not -- the panel's input of whether or not we have gone too far on some of these markers, on all of them, or what.

I do not think there has been any inconsistency on

the Agency's part in enforcing what is -- or dealing with what is in the guidance document. If anything, it is the fact that we have been consistent across all the markers and probably exerting an overkill on some, whereas holding all of them to a fairly high level proof at the present time. Steve, would you want to comment on that?

DR. GUTMAN: Yes, I do think that there is some confusion here, and that as a result of the input, we either need to perhaps change the way we interpret this guidance document or modify this guidance document. And when I either brilliantly or idiotically was trying to present the case to this panel for the down classification, that at the time seemed like a very novel idea and one which we were not really quite sure the panel would buy or not, although it is now met with accolades.

It was the suggestion that we were going to lower the administrative standard, but not to change the clinical standard. And it may be my fault, in part, because my interpretation of the clinical standard was that we were going beyond the analytical part -- and I may have read the guidance -- or folks may have read the guidance too fast and assumed that we were not going to make a distinction between the next PSA and the next PUK, or whatever somebody thinks

of that is fantastic and novel and has got an antigen or whatever.

It really is important to us. We really are coming to the panel with an open mind about looking for appropriate thresholds. Our challenge in this new regulatory paradigm is to find appropriate minimum thresholds, not to ask for data simply for the sake of data, but to ask for data because it makes sense and it enriches the label and it provides information that is useful to people who are going to buy the product.

We do think that some of the ideas, the difficulties -- I am not a statistician and we were unfortunate in that we do not have the statistician present, but the issue is real and it is one that the statistician will not be able to deal with. You take small populations in different designs and different intended uses, and you will get an estimation of performance but it is not going to be rocket science, and we probably do not want to start asking the manufacturers for -- you know, for something like the first row(?) study on CEA.

It really is important to us, and at some point, I will not let you out of the room without going through -- and you have all actually provided some, I think

fascinating, actually beguiling insights into this, to encourage us or to discourage us from crossing this threshold and moving to a more analytically-based characterization for either high risk or low risk or for well-characterized or poorly-characterized, or for ones where you know you are looking at the same epitope or different epitopes. It really is important for us because we take your opinion very seriously as a way of grounding what we are going to do.

We have two choices as a result of your input. One is, we certainly are not going to do nothing. We will either follow the guidance document as it is written, which is to make this analytical cut or clarify it in some cases, or to change the guidance document if you feel that there is some clinical data set that is necessary, and try and clarify what that clinical data set ought to be.

DR. LADOULIS: Dr. Jordan?

DR. JORDAN: Something that Dr. Homburger said that rings true. I think we should look at outcome more to the low risk versus high risk, and to me there are differences. But just trying to frame -- get this in my mind, if a me too comes along, is there a difference in insensitivity and specificity that it has to have from the

original? How much of a range does it differ?

DR. GUTMAN: Well, we do not have any concrete standards. It is one of the things that has always bedeviled us, is trying to know when something actually is equivalent, and our general response is to try and make sure it is just not alien in outer space, so that it is in at least the same ballpark, and that is one of our hang-ups about the labeling. It may be an obsessive and self-deluding hang-up, in that we may be generating information that is not valuable, but one of the reasons we have always requested these small data sets, is the notion that we do have something that we can hang a label on so you can buy it and think you have an approximation of sensitivity or specificity.

If that is not true, and there is some other alternative, well, that is great. And if there is some truth to that, then we do not want to abandon that.

DR. JORDAN: Well, that is scary for me. Because on the one hand, if you are going to give me a me too, if it is a high risk, I would certainly want to have at least a 99% but no less than 95% deviation of sensitivity and specificity.

At the same time, I have to go back to Dr.

McCaskill-Steven's point earlier, what good is looking at what the original test has done, if the group that was tested is very different, or is basically alien to what we are doing now, too? So, I mean, you have two different areas here that have to be dealt with beforehand, and particularly for a high risk.

For a high risk, I want to make sure that the original group that was tested was broad enough that this me too is encompassing. At the same time, this me too is falling within at least a 95% sensitivity and specificity of it. Otherwise, it is not a me too.

Now, for a low risk, there may be a wider range, but I think we get to the definition of, what do we call a high risk versus low risk?

DR. RABINOWITZ: I wonder if I could make a couple of comments just real quickly. I think the reason we do clinical trials with FDA products is to look for the unanticipated, the unknown that might arise. For example, if one were to characterize the antibody, as Dr. Homburger suggested, you might not realize that -- you might not know enough about the metabolic pathways of the analyte, so that you may have a perfect knowledge of the epitope of PSA, and know how the antibody, the monoclonal antibody you designed

works in an analytic setting, but you may not realize that it becomes bound by some protein and that covers the epitope, or something like that.

I think the main reason that clinical studies are done is to look for what you do not know, even though you characterize it. But I personally think that the more we could be like the generic drug regulation, the happier the clinician would be, because there is an expectation as a clinician, I think, that a PSA is a PSA; a glucose is a glucose, and yet, many of these analytes are not as easy to characterize as a drug, and the kinetics are not as established.

I wonder if there could be a few comments on what sort of analytic data would be important to add to the guidance?

DR. MCCASKILL-STEVENS: I worried about the fact that if we were just doing the me too setting, you are just comparing the one test to the other test. If we do not look at a target population, then for instance, again, the new me too antigen might be more sensitive than the old antigen, and it may look like it is off, because it will not be as sensitive. If we take the other one as the gold standard, then it may look like it is off, when in fact in a special

population, it may be better.

DR. LADOULIS: I think to amplify that, we have noticed that in the past, that there have been differences even between the spiked samples where the antigen is put into the appropriate matrix, and I assume that, meaning about the characterization, you include not only the antigen and the antibody, but the matrix. Which is why we have interference studies, etcetera.

The best gold standard for a measurement of an analyte is a native sample from a patient with that analyte, then tested by an approved device, and the candidate device for approval. Even reference bank samples have problems that industry and everyone recognizes are problems.

If it is required to do due diligence and therefore to evaluate a device, that you have fresh sample. Then it automatically implies that you have some current targeted population. However small in number. So, it seems that for every me too product, there need to be some fresh samples of the patient, and that might be answered by having a small number of samples, but also perhaps as Dr. Reynoso pointed out, having short term serial parallel, tandem studies, both by the approved device as well as the candidate, to assure that not only is it in the same matrix,

fresh sample native analyte, are the same things being measured, but no interference occurs even in the same patient over time by the approved device, or discordance between the approved device and the candidate device.

I think there have to be some kind of targeted clinical populations in order to overcome all the concerns about differences between bank samples and fresh samples, and the differences between matrix affects and current patient setting where there are new types of therapies being introduced, as was mentioned before.

DR. TAUBE: But you are still constraining the study a little bit more than what is required in a full clinical study. I mean, if you had a small set -- you are not asking, the way you just said it, that the manufacturer redemonstrate the entire clinical claim.

DR. LADOULIS: No.

DR. TAUBE: You are only saying that there should be a comparison with a set of --

DR. LADOULIS: Just performance.

DR. TAUBE: -- fresh samples. I thought the data that Peter put up earlier was interesting that you can show concordance of tests, then if you had only tested the one test for its sensitivity, it would have looked very bad,

compared to the reported sensitivity of the other tests, but when testing the two devices side by side on the same set of samples, they showed that the sensitivity and specificity were in the same order of magnitude in those samples, which was different from the original report of the originally-submitted device. So, perhaps the study has to have the true side by side comparison with a set of fresh samples, which is different from what was originally submitted. And that, in addition to more information on the actual antigen-antibody interaction --

DR. LADOULIS: Yes, all I am implying is to validate substantial equivalence in performance, not the clinical claims for the use of that device.

The other point I just wanted to amplify is something in terms of specifics. When PSA came before the panel, there was rigorous review of some of the original data. It was discovered that in fact, the population of normals used for the evaluation had a mean age of 50.

Secondly, it was finally discovered in the process of just for the questioning and interrogation and evaluation of original data, that the population of normals included 95% white people, okay?

There are instances in which there are approved

devices for claims in the market, and they get modified in the future because they were not maybe the best populations to study. We should not penalize necessary subsequent candidate devices because they have not chosen the populations which we thought, you know, at one time were the right populations or not the appropriate populations, or treatments change. But as long as the performance characteristics are the same, if they choose a smaller number of a normal population which are age 70 instead of 50, and if they are treated in a certain way, and if they include all the special populations as well as some other majority populations, it does not mitigate the value of this me too device, and does not place an undo burden on the sponsor, I think, to get this device approved, as long as side by side, in a clinical test, the approved device and the candidate device perform the same.

DR. REYNOSO: I wonder what my fellow panel members would say about the possibility of accepting that there may well be a number of me too tests for which enough is known in terms of the immunochemistry, the physical chemistry, the antibody and antigen characterization, for which substantial analytical equivalency may be assumed. And in which such group of tests can they be defined. And

then we can deal with the old group.

If we could agree on that, maybe we could can to grips with the question of how to define the two. But in my mind, I think that there is a group of tests -- me too tests -- in which analytical equivalency may be sufficient and I think that we could agree on that. Not for every me too test, but certainly for some. Then we could deal with what it will require of the other ones.

DR. LADOULIS: Well, I would disagree with -- if I understand you properly -- is that analytical equivalence would have to be defined in terms of an analysis of a fresh sample, and not a synthetic substitute. Because, even measuring glucose, spiked glucose sample compared to a patient's --

DR. REYNOSO: Define the terms in which you can -- [simultaneous discussion] --

DR. LADOULIS: Well, maybe we have covered enough ground, but the FDA has got sufficient information --

DR. MAXIM: I am ready to move onto question number two. I would like a little bit of clarification -- when you referring to a fresh sample, do you in fact mean, freshly obtained sample for that particular test, or are you talking about measuring the analyte in its intended matrix,

such as having serum or having plasma, rather than spiked samples?

We can establish based on ancillary testing that a particular analyte may be stable in the freezer for extended periods of time, and come back and look at banks of retrospective samples, serum samples, on some of these patients to get this information, rather than going out and initiating studies where you are doing fresh draws in a prospective manner on these individuals.

DR. LADOULIS: Well, I guess an immuno-chemist, Dr. Homburger and I and Dr. Reynoso might want to comment on it. Anybody want to make some comments about it? I think that my initial reaction to this was the same criteria has have applied to the original applicant for the first introduction, which have often been frozen samples that have not been obtained fresh from a patient.

In certain instances it was necessary to treat the samples, or analyze samples that were not frozen or banked, but not frozen and thawed and re-thawed samples that have been banked, or that have samples that have certain stability characteristics that we know, that tolerate a single freezing, might be satisfactory. But, I do not know. It depends on what the analyte is. Any comments?

DR. HORTIN: Yes, I think it is possible to define a set of circumstances under which you can maintain a bank of biological samples, be it serum or whatever, that can be used to look at different analytic methods that purport to measure exactly the same thing, to see if they measure exactly the same thing.

Many of these methods are sufficiently variable that they do not measure -- if it appears that there is a fair amount of scatter, it is probably because of the measurement, not the sample, anyway, but I do think it is possible for many analytes -- many of those included -- to define some conditions under which it would be reasonable to look at bringing together a bank of biological samples.

DR. REYNOSO: I have no further comments.

Agenda Item: Question Number Two

MR. MAXIM: I am not sure this is considerably different from what we have just been discussing, but essentially the question is, is there a point where there is no longer a need to review clinical performance, rather than just a side by side comparison of the approved, or cleared device, with the candidate device, on a given number of serum specimens.

If we look at, perhaps you could extend the evaluation to the distribution of the marker in normal individuals, which correlate to the patient population you are going to be studying. A panel of benign diseases which you would expect to be involved, or also express the antigen and the serum, and different malignancies, as well perhaps the different classes -- Stage I, II and III of a particular patient, so you can see the expression or the distribution of the antigen across the various samples.

This would be single point determinations in a significant number of serum samples, or include the performance and, as we have been talking about, the targeted patient population? If the tumor marker says that it evaluates patients for the recurrence of prostate cancer, do we need a select number of serial samples from prostate cancer patients to be sure that the marker does that as well as the candidate device?

It is a graded degree, I would say, of clinical samples, and we are wondering if the panel had any comments on this?

DR. HORTIN: I have a comment. I think if you reach the point where you are comfortable that you want to make an equivalency determination based upon analytical

comparison, and that implies that you are comfortable with the considerations of differences in patient populations and treatments and whatever other considerations one wants to consider. Then it is simply a statistical question as to how many points and how many samples of each type you need, in order to show equivalence. And that is based upon the degree of variability of the analytical methods that you are comparing, and the known heterogeneity of the analytes and that sort of thing.

DR. GUTMAN: But I can point and ask for input, that the two dots really have very different implications in terms of study design, because the first is asking for a cross-sectional study in, as Peter defined it, a variety of different settings, and you can decide how many and what settings and what statistical numbers to apply.

The second implies at least the possibility of an actual -- not outcome, because we do not really do long term outcome -- but, certainly, a prospective study predicting what you expect the tumor marker to be doing in the patient set. So they really --

A and B are really quite different, and I would suspect -- and the manufacturer is welcome to get up here and agree or disagree -- that A represents a change in the

threshold we have now, and a lower threshold, and it would be interesting to know if for a me too marker, the panel was comfortable with that as a threshold.

DR. LADOULIS: I will comment on one thing, that is that, performance equivalence is what I think should be a major criterion, and for that the samples must be assessed at -- or near -- the threshold that is clinically determined to be significant by a previous device. So, that is one consideration. And the statistical considerations in terms of a minimal number that is required can be determined, based on the performance characteristics of the candidate device.

As far as the longitudinal studies, I think that is beyond the performance characteristic requirement, or criterion requirements, and they do not need to be undertaken by a sponsor for that purpose, because that addresses the justification of the need for that assay, which has already been established. Clinical need and its utility and managing implications.

Here, I think it is a distinction of performance. Does that answer -- that answers your question, I think.

DR. TAUBE: So you are saying that it should be the second bullet, and not the first?

DR. LADOULIS: Well, except to the extent that the first bullet refers to some interfering conditions, and I think you already have, still, in the guidance document, a requirement for some performance testing to be done in order to evaluate the interference due to different drugs, than maybe a hemoglobin or a bilirubin, or some other alterations of the patient matrix. Is that right?

DR. GUTMAN: Yes, but you can see interfering substances in the cross-sectional study, so the issue is, if you wanted to require -- see, there are several things on the table here, and I really am trying to make sure I am following.

DR. LADOULIS: As long as they are limited to performance issue, performance criteria.

DR. GUTMAN: Okay, because -- [simultaneous discussion] --

DR. LADOULIS: And that might require some other conditions of some patient's sera for that purpose.

DR. GUTMAN: Okay, but are you favoring the first dot, the second dot, both dots --

DR. LADOULIS: I do not want to -- can we blow the dots and take the focus on the issue of what is the criterion? The criterion is -- I do not want to necessarily

be confined just to those dots.

DR. GUTMAN: Okay, well, you are allowed extra dots, or a new page.

DR. LADOULIS: Or wipe out the dots. Or connect the dots.

DR. GUTMAN: It seemed to me -- I think it was you, what I thought I heard -- and again, I think this makes a difference both in terms of our review and also in terms of the threshold you are making for the manufacturers.

It seemed to me when you were talking about prospective studies before, you were talking about looking at longitudinal data, comparing two analytes, but not necessarily with a clinical outcome. More to see one analyte versus the analyte, and whether they were changing over time in tandem. Did I interpret that correctly?

DR. GUTMAN: Well, one of the suggestions from the panel was about doing longitudinal sequence, or serial determinations on the same patient in order to evaluate some idiosyncratic performance characteristics of one device against a previously approved one. That might not be necessarily what you want to adopt. But if you used just the one single point determinations, and used a targeted population, however small, it would only be for the purposes

-- that population would only be selected for the purposes of defining the performance against an approved device, not for the clinical purpose of diagnostic or management as to what the threshold or of rises and fall in the analyte might indicate in terms of clinical outcome.

DR. HORTIN: I think the need for the clinical studies really publicly relates to a degree of agreement. If you have virtually perfect analytical agreement, generally you are not going to turn up much in the way of a difference in the clinical studies.

Again, that may be a problem somewhat in terms of defining exactly what is sufficient, in terms of precise analytical agreement. You get back to kind of that problem a little bit. And oftentimes those relate to the issues of having well-defined antigens and antibodies that we have talked about. If you use the same antigens and antibodies you will generally get very high agreement.

Some of these issues are relating back to some things that we have discussed before, but I think to some degree, if you have extremely close analytical performance, you are probably not going to turn up anything different by doing the clinical study. The classification is basically going to be the same and it is basically going to be a

redundant process, and oftentimes I think you may find a study in kind of intermediate ranges where you may have a few outliers, and I suppose if you wanted to be most efficient, you could perhaps concentrate the clinical evaluation in terms of evaluation on the outliers, by trying to assess basically in terms of which one is giving you the better clinical information or not.

That would be perhaps a way to improve the efficiency. Instead of doing extensive clinical study, if you can identify by statistical basis, there are means of evaluating whether a point is an outlier or not, and evaluate clinically those that would reduce the size of the population that would have to be evaluated a lot, and might be a way to improve the efficiency of doing the evaluation.

DR. AMMIRATI: That sounds very appealing on its surface and also I would like to play a little devil's advocate. The negative of that is once you have identified two or three outliers and said, see, my test actually compares better with the clinical outcome, or what you see symptomatically. Persons will come back and say, well that is anecdotal. It is not enough. It is only three people. I think there is some concern about spending too much time on outliers just for that reason, that they are hard to

support one way or the other.

A couple of more comments. One is that, if we think about what we are really trying to show with substantial equivalence now that there is reclassification to Class II, and I think that really should be maybe at the center of a lot of these discussions. Granted, these are newly-reclassified and they are tumor markers, but if they are Class II and they need to be substantially equivalent to the predicate, then if you look at all of the other Class II analytes our there, then basically we do look at analytical performance, and a lot of the issues we are talking about.

Coming back to testing longitudinally, I think that the whole issue was really marked very clearly this morning in Peter's example. I think one of the graphs had Test A that came to market with 58% clinical sensitivity. That is really good for a tumor marker, which means that almost 60% of the time, whatever threshold was chosen, that that person was positive, to use those words, for the clinical condition.

In another study, that fell to 37. I assume that the protocol designs were somewhat similar, because that is how we go about setting up our studies, as somewhat similar to ones before, and I almost wonder if you came to panel the

first time with a 37%, would there have been the same outcome, because that is really different.

From the regulatory point of view, I am almost thinking, now that this test has shown 37% clinical sensitivity, do we begin to do an adulteration of his branding issues? This is so far away from the 58% that is in the labeling now, that we have to think about that.

Clearly, just the luck of the draw, depending on who your patients are, how sick they are, because generally these studies have the same number of patients. Those are pretty well controlled by FDA and industry. You can get very different clinical outcomes. So maybe we need to rely more on the analytical aspects of what we are measuring, because the clinical aspects for target populations give us very different pictures, if you look at the difference --

We are at a little bit of a disadvantage, because the people who order the tests are not the same people who purchase the tests, and we talked a little about this at lunch. If you look at the laboratory, you make a decision based on choosing one tumor marker, based on a lot of things, not the least of which is economics, which we know is separate from this panel to begin with.

I am wondering from maybe the clinicians on the

panel, if you knew that this test only had 37% sensitivity, would that change your opinion, as opposed to it had 58? And those are the problems we run into when I think we rely on those kinds of clinical studies.

DR. LADOULIS: I think that leads into a question I think that we addressed I think before, and that is, whether to use the same descriptive statistics of sensitivity and specificity just in the broad terms, as sufficient? And whether these should be actually qualified or use ROC curves or some other kind of criteria, because as you pointed out, depending on the study and about the selective bias that might be introduced in a population, as specificity and sensitivity are percentages, and they do not have the same kind of clinical value to the end users. Or they might not over time. Or else they might not -- there might be such variation that 58% is no different from 37%.

DR. AMMIRATI: Right. Right. That is what happens.

DR. LADOULIS: And yet, it might be promoted as somehow -- or inappropriately.

DR. AMMIRATI: Better. Right.

DR. LADOULIS: So the question I think I asked Max or someone, if there was perhaps some other alternatives

that the Agency might consider in terms of evaluating performance, statistically. Comparing a candidate device with an approved device.

DR. HORTIN: The discussion that I have heard around the table suggests that we are struggling to write a means of defining equivalency, when we are dealing with analytical methods that are --

When we are dealing with analytical methods that are inherently very variable. They are variable in terms of the antigenic sites that are detected, the antibodies and the methods themselves, and their application in a clinical setting, in which the patients are highly variable.

There needs to be some sort of a gold standard for comparison here. Now, if we want to take the tack that the lowest common denominator is analytical equivalence, we have to ask for a lot of detailed information about what is being measured, how it is being measured, probably more than anybody is providing right now, certainly more than they are putting on the labels of their products.

It suggests that there needs to be -- the notion of a serum bank came up. A serum bank is not a bad way of doing it, provided the serum bank is rigorously maintained, you know, and perhaps we could tell you how to do that, or

scientists could tell you how to do that, obviously no one would freeze and thaw things multiple times, etcetera.

I do not think it is realistic to run out and recruit 300 patients every time you want to bring a small thing to market, and try and follow these patients for three to seven years, that is what those folks are saying, and that seems to me to be quite a reasonable position.

The idea of a serum bank is not a bad idea, but that raises some other issues, you know? Whose serum bank is it? How is that bank maintained? What input does this group have into describing what goes into that bank and how it is maintained, and how the alloquots(?) are handled and doled out and all that sort of thing?

Now, I do not know whether that is an acceptable alternative, but when you are looking at trying to control two highly variable things, one the analytical side and the other the clinical said, and then show that they are equivalent, you need something in between to hang them both on, and the notion of a serum bank is not a bad one.

DR. REYNOSO: Yes, the question of a serum bank is interesting. Speaking to the subject of a serum bank, that may very well be a very viable alternative, and many, many years ago, we set up serum banks in the National Cancer

Institute, the Advisory Panel there, and actually wrote the Request for Proposals and people, including the Mayo Clinic and others, submitted proposals for developing serum banks that met the specifications of the National Cancer Institute.

It can be done that way, so there is a way doing it that way. And anyway, may I speak for just two more minutes? Other than the serum banks, which I think is a good idea, and they can be done well. It has been shown that they can be done well. To the points on the slide, I think the answer is yes to both points.

I think that there is a point where there is no longer a need to do anything beyond clinical -- beyond analytical equivalency. And at what -- and both thoughts -- And if so, at what point is analytical data in comparison to a legally marketed device sufficient?

I think that that could be defined, that could be defined in terms of experience of the FDA, in terms of how well defined the test is, the original test. How well known the immunochemistry and histochemistry(?) are, and simple questions of just, how similar are the antibodies? Are similar is the antigen?

One defines a set of circumstances, or variables,

one could say, yes, we are at the point at which any other me too test coming along does not need further clinical studies, or a few limited ones. And analytical equivalency is sufficient. So, I am answering yes to both under some defined circumstances.

DR. TAUBE: I would just like to address the serum bank issue. From the point of view of the National Cancer Institute and the program that held that enormous serum bank over a period of 20 -- over 20 years. I was not in charge of it during that whole time, but I can say that, in theory it sounds like motherhood and apple pie and it should be wonderful, but in practice, we chose with advice from the outside, to finally close it down, for the following reasons.

It is either exorbitantly expensive, if you keep renewing the specimens so that you do not have them sitting there for 20 years. We had tremendous quality control on the alloquoting(?) and testing over time for things like CEA that are stable over a long period of time, and for some analytes that we expected were not stable and so there were alloquots that were taken out and we did time core(?) studies, and so on.

The problem is, that it still a single draw. At a

time we had normals, benign, other diseases and -- from various tumor sites -- patients who were diagnosed with cancer in various tumor sites. It turns out you rarely have in your bank, no matter how well designed it is, exactly what people need at any given moment in time.

You always have this problem of samples getting very old; of diagnostic criteria changing, so that originally most of the samples that were collected were collected according to the old Sears staging which is not very relevant in today's situation. And so, it is very hard to maintain a serum bank over a long period of time.

There are some other serum banks around that are being collected now. The Gynecological Oncology Group has a very good serum bank on ovarian cancer patients. But these are designed not for perpetuity, which you would sort of have to have if you were going to address the situation of having standards around that people are going to use every time they wanted to bring in a new device. It just -- everybody always wants a serum bank, and I am just telling you from experience the answer.

DR. LADOULIS: If there were a limit to the number of patients that were required, and the number of samples, then it is possible that in practical terms, maybe fresh

collections over a short period of time for single point determinations is a more reasonable alternative than the relying on a bank. Is that what you are implying?

Are there any other comments or any other questions you have from the panel with regard to this point? Peter?

DR. MAXIM: I was just wondering if anybody else wanted to speak to question number two.

DR. LADOULIS: Any other comments from any of the panel on this point, on either the kinds of samples to be collected, or to be obtained or to be used, or how the data should be characterized?

Agenda Item: Question Number Three

DR. MAXIM: Actually, we may have just addressed this one, also. If the FDA were to reduce clinical -- the panel is about one step ahead of the questions. If the FDA were to reduce clinical data requirements, which of the following do you consider to be valid prerequisites?

Standardization of the assay. We know standardization if proceeding with at least a PSA marker, perhaps a couple of the other ones.

Comparison to well-established reference methods, which are very, very limited right now.

Performance with an accepted reference, and the serum panel that Mr. Zalesky mentioned this morning and we recently discussed I think would be one of those accepted references.

Or, the Agency dealing with the reduced clinical data requirements as part of the labeling, to state that the clinical performance of this assay has only been established under a particular set of circumstances, or has not been evaluated at all to the expected values for particular cross-sectional populations that they may look at, or whatever we go to, but not actually state that we have looked at the clinical parameters as defined by the predicate devices. Any comments on these?

DR. LADOULIS: Anybody think we have addressed this already? Maybe we have probably covered them. Any comments from any of the staff?

DR. MAXIM: The last question dealt with one of the issues that Mr. Zalesky had brought up this morning and that listed the various impediments and the ability to obtain clinical data for tumor markers. He listed the availability of first of a kind markers. Availability of home brew assays. The small study sizes that are necessitated by many of the tumor marker studies. A study

duration may exceed the standard of care, and studies are impacted by treatment and therapy.

Does the panel feel these issues preclude clinical studies for all tumor markers, other than the new first of a kind, or can these limitations be overcome, and if so, how? And again, I think we probably alluded to this this afternoon, also, but if there is any additional comments or anybody wants to speak to any one aspect of this, the Agency would be interested in hearing the comments.

DR. LADOULIS: I think one -- do you want to address something about the availability of patients?

DR. TAUBE: I thought some of his points were valid. That once you have a test out there that people are using, it becomes difficult sometimes to set up a trial, but -- to try the same kind of test, and it is hard to get people to -- physicians to enter patients and so on, and in the climate of managed care as well it is also difficult.

That assumes that you are talking about some large enterprise. I think to get a small number of specimens to do the kinds of studies that we have talked about in terms of equivalence, may not be so difficult.

DR. LADOULIS: I agree. I wonder if there are any other comments about the comment, or the criticism that was

laid about availability of particular types of patients, such as, you know, patients with unusual tumors, for example?

DR. KEMENY: I do not know if you have mentioned or if I missed it, but have we discussed -- when we are talking about equivalence, what are we going to accept as equivalence? I mean, is it going to be 100% equivalence or 90%, or 80%?

DR. MAXIM: No, we have not discussed that yet. Substantial equivalence, technically, does not relate to degree of comparability, or agreement within two assays to a particular level. It is simply that the device performs in a manner comparable to something that is already in the market.

That is something we would be very interested in hearing if you had any comments on that, whether we set limitations on it. It is not something actually that we have done with these markers in the past, as far as within particular limits.

DR. LADOULIS: Can you explain that? Does that mean that -- I mean, to what extent, then, does it mean when an approved device is put out there for a particular marker, that its value determination for a particular tumor marker

is substantially equivalent to a value of another marker? Is that not implied?

If it is 10 nanogram/ml by a legally approved device, plus or minus a certain standard deviation --

DR. MAXIM: That is in its label. Its performance is -- if it is a truly quantitative type of assay, then what you will get on the labeling will be the fact that, when it measures ten nanogram/ml, it is in line with its calibrators, and that it performs in a manner --

DR. LADOULIS: But it is substantially equivalent to other available standard assays?

DR. MAXIM: Right.

DR. LADOULIS: Previously approved at this same --

DR. MAXIM: Right.

DR. KEMENY: But the substantially is what we have not defined. And I -- so, if it is not defined, then I would say that, since this is a kind of -- if we are allowing the me too thing to happen without clinical testing, then we are only talking about equivalence to laboratory -- in vitro testing. Then I would think it would have to be extremely high.

DR. MAXIM: Concordance with another assay.

DR. KEMENY: Yes.

DR. MAXIM: That is what we alluded to this morning. If we go back to strict analytical type of parameters. That is not to say that they do not -- that we do not do that right now. We do not look at the analytical performance of the tests side by side, but there are acceptable limits for each assay.

The point really is for its intended use. The performance and the clinical population is also part of the substantial equivalence. And that is where it gets a little harder to put real numbers on the assays.

DR. LADOULIS: Well, let's take an example of normal cut-off level for a particular analyte. A legally accepted device has been established and the legal cut-off limit has been established. Now another device is being proposed by a sponsor. And they normally will have to do some comparison studies, and there will develop performance.

It is very unlikely that a value from a candidate device for a set of samples will be exactly the same as that for a legally approved device. So if the cut-off was five units, or five nanogram, or whatever, and that is the clinically useful cut-off being used in the management of patients for a particular device, and now somebody introduces another device, and there is some systematic

bias, so that there is a 5% difference higher by the new device, compared to the old. Does that mean that this device should be labeled as having a cut-off now which is different from the previously clinically accepted cut-off for other devices that are previously approved?

That is, that a normal patient will now be 5.2 instead of 5.

DR. TAUBE: But when you compare the performance that each device can have a different cut point. The cut point has to be evaluated by the ROC curves and so on. They can establish a cut point such that, above a certain point you have a certain sensitivity and below and so on. So that, the population is divided in the same way as the other test. And then, when you look at the specimens, and Test A gives it positive and Test B gives it positive, that performance is equivalent, even if the cut point that is used for the particular test is different. Isn't that correct?

DR. KEMENY: I think there are two issues. The one issue is that, you are looking at whether they have the same values, the two tests have the same values. That is what you are talking about. What I was talking about is whether they have the same sensitivity and specificity,

which clinically is more important, because I mean, the test has to decide what their values are.

I mean, if it is CEA, for instance, if it is five or ten, I mean, they have to say what their normal value is. But more important is, is the specificity and sensitivity the same?

DR. LADOULIS: Right. And those two are related, because now it is the sensitivity and specificity at the clinical --

DR. KEMENY: Not necessarily.

DR. LADOULIS: -- at the clinically useful level. And it has to do with the operating characteristics of one device, as it is used, over that clinical range, versus another.

DR. KEMENY: I do not see that.

DR. MAXIM: I think basically we are dealing with two separate issues, and again, as Dr. Taube pointed out, that if you have two devices coming to the marketplace, they would establish a cut-off for that specific product with regards to their clinical parameters; clinical sensitivity and specificity.

The cut point may in fact be different. Device A may recommend that the cut-off to determine the sensitivity

and specificity they selected from the ROC curve or whatever form of analysis they used, is 37.9, or 40 units/ml.

The second device came along and did a similar study. They would establish their sensitivity and specificity, much like we looked at this morning, with a cut-off of perhaps 35 units, or 32 units. So with that device in that particular population with that intended use, we would have two separate cut-off points to determine that you would have this clinical sensitivity and specificity associated with it.

The other type of cut point would be something, for instance, with a clinical ramification as specified by PSA at four nanogram/ml. In that case, we would be looking at the manufacturers as they came in, to establish, based on population studies and looking at the performance of the device, that they in fact could support their cut-off, and these populations would be four nanogram/ml.

That is a much more clinically related cut-off than is --

DR. LADOULIS: They could not establish clinical sensitivity or specificity unless they did a target population.

PARTICIPANT: Right.

DR. MAXIM: Well, not with sensitivity and specificity, but the -- [simultaneous discussion] --

DR. LADOULIS: But you can get a response -- you can get an operative curve characteristic for the device, and that would have to be compared to the characteristic for the legal device at the range below and above the clinically useful threshold. So that is what you can do. But you are not comparing it with outcomes, if you do not have every subsequent manufacturer comparing it with outcomes of patients.

DR. KEMENY: No, because you could do it -- for instance, we could take -- not this, but like alkine(?) phosphatase which, you know, is different in practically every hospital you go to. But they have normal ranges.

I mean, there would be normal -- you know, normal ranges for each test. And in other words, if they look at it against ten tests of the antigens that we know, they would figure out what their range is, but then they still might -- there might not be concordance. For instance, they may have some tests where, let's say, a PSA on the -- you know, the tests they are looking at is ten, and they come out -- theirs is zero. I mean, in other words, it is like, off. No matter how they look at it.

There has to be concordance or equivalency with the specificity and sensitivity for the test to be equivalent. It does not have to be the same number, just as long as --

DR. LADOULIS: No, certainly for normal patients.

DR. KEMENY: Well, for normal, they have to say what their normal range is. It does not have to be the same number, though.

DR. LADOULIS: Right, for normals. Not the disease-affected individuals.

DR. KEMENY: Right, but also -- normal and diseased.

DR. LADOULIS: But I guess the issue I am saying I am concerned about is, say, for patients who clearly have evidence of disease, and have a level, for example, a PSA of ten. Now, what does this -- for the legally marketed device, that represents a risk -- predicted fact -- level, which is very high. If the patient is 50 years old, let's say. But for this new candidate device being sponsored, does that -- a level of ten in that device, add the same specificity and the same sensitivity, for the patient with disease? Say, predicted value.

DR. KEMENY: Right.

DR. JORDAN: The numbers would be 75, so long as you when you are look and compare them side by side, these mean disease, these mean no disease, and there is no overlapping.

DR. TAUBE: Right.

DR. JORDAN: I do not care what the numbers are. But particularly in a high risk situation, they should be very close. There should be a very small deviation from one to the other, as I said awhile ago -- [simultaneous discussion] --

DR. TAUBE: But they do not --

DR. JORDAN: -- same number. They are measuring different things.

DR. TAUBE: But you only need a small set of samples to do that test on, so long as you are doing the side by side test of the --

DR. LADOULIS: But at the extremes, there is no question. The cut-off level is because there is a level of probability, an increased risk of disease. Five nanogram/ml represents something, you know, of significance perhaps to some population, certainly not others. And it may be significant to some urologists, but someone else who is experienced with his population it does not mean too much --

We are concerned at the debates that we had about PSA, as to what would be the impact of having a particular cut-off on unnecessary testing of patients with benign prostatitis, etcetera, in the elderly population? A certain cut-off was chosen. And it was obvious that there were age-specific normal ranges, and so the problem is that this critical threshold where there is a suggestion that some other clinical decision needs to be made about further testing, for example, an ultrasound or biopsy or something like that.

This had to do with more or less diagnosis than really monitoring, but the same applies to patients who have been treated, have no evidence of disease and are not being monitored. At what level does the rise or the value of a tumor marker now have clinical significance to require an intervention?

We are talking about not extremes, which there is no question about a concordance but, clinically useful monitoring. How can we compare now, what is a substantially equivalent test, one that is -- [simultaneous discussion] --

DR. JORDAN: That is for the manufacturer to decide when he comes to FDA, I assume. It is not for the FDA to draw [off microphone] -- If the numbers do not fit,

he does not get the approval for a me too test. I mean, to be a me too, that means they have to be, again, falling within that high range. And if it does not do that, if he comes up with a number that does not match, then he does not get a me too.

DR. HOMBURGER: To go back to your example, if you came to me with your PSA test and you said the cut-off is 75, and the one you are comparing it to, the cut-off is four, and the sensitivity in your group of 50 patients was 85%, and the sensitivity in the group where it is four was 85%, I would not feel very good about that. Because I would think your think should be measuring the same thing as the other one that is out there, and I would be asking myself the question, boy, there is all that difference between the two. Now, is that an artifact of the small patient sample? Is there something else in your standard that is causing yours to read high, or is mine calibrated wrong? What is wrong here?

It gets back to this interplay between descriptive data about the method, and descriptive data about the patients, before you can make any decision, really, about equivalent. And the more you know about the method, the less you need to know about the patients.

If you have a huge sample of patients, and you do not know much about the method, you do not need to know much about the method, either, because if you study enough patients -- you study 20,000 patients, and yours and mine work the same way, I think most reasonable people would accept their equivalent without knowing much about what is going on analytically.

I guess this is a balancing act. The more you know about what you are measuring, the less rigorously you have to test it in the clinical environment, and the less you know about what you are measuring, the more rigorously you have to test it in the clinical environment.

To make a comment about that question, I would raise the issue, is that really our concern? Now, if I were a manufacturer, and there was a $50 billion market out there for something, as opposed to a $3 million market out there for something, the degree of, how do we say it -- impediment -- is proportional to, I will jump over a lot of impediments to get a big share of the big market, whereas I will put the thing on the back burner entirely. But I do not know that any of those considerations are really relevant to the basic issue of equivalency and being able to show that.

DR. RABINOWITZ: I just want to make a comment,

and this is my own opinion, because I have never worked in the Center for Devices, or for the Center for Drug Evaluation Research, but it is my understanding that for generic drugs, the standard is, that once the innovator drug no longer has its patent and so forth, and then the generic drug manufacturers come in. They do not have to prove the use of the drug, but they have to prove comparability, and its bio-equivalence, usually done in a certain number of volunteer normal patients, about 20, I think, and it is having a bio-equivalent, plus or minus a figure, and I think it is 20%.

If the innovator drug is 100 units of drug --

PARTICIPANT: Plus or minus 25 or something --

DR. RABINOWITZ: Plus or minus 20% is what the threshold is for the generic, and I think Dr. Homburger's suggestion is perhaps, if you know enough about the analytes and the epitopes -- or the analytes and the antibodies -- comparable to what is known about drug structure, and then could make the analytic performance comparable to the bio-equivalence, that would be something that would enable the clinician to have some assurance that the four of one test was comparable to the four analytic value of the other test.

DR. REYNOSO: I think some of those points were

well made this morning that are well made now. My comment is that, many of these points do not preclude the performance of clinical studies, they just make them difficult or perhaps economically unfeasible.

The real question is, are they necessary for me too tests? I mean, they are more difficult if the feeling is that they are necessary, well then they are going to have to be done, however difficult. The question is, are they necessary? And I think part of the point that I hear is that they may not be necessary in all cases. They may be necessary in some.

DR. LADOULIS: Or they may be even necessary only to a limited extent.

DR. REYNOSO: Right. Or to a limited extent.

DR. LADOULIS: Is it the sense of the panel -- do I hear correctly that it is not reasonable to expect that no clinical studies should be done, but that some clinical evaluation has to be done, is that what I hear? Some for assays.

DR. KEMENY: I think most of us feel that in some cases -- and that needs to be stratified -- that in some cases equivalent studies without clinical studies can be done. And in other cases, a clinical correlation might be

necessary.

DR. REYNOSO: And one way of working the correlation would be, can this guidance document be left alone, or does it need modification? And I guess it needs modification, perhaps to reflect some of the things that the sponsor said, and some of the things that have been said around this table.

DR. LADOULIS: That is why we are here, I think, is the Agency wants to know whether or not the guidance document is -- do you want any more information than what you have already heard, or should we summarize it? Do you want to summarize what you heard?

DR. MAXIM: No, basically I think what we were looking for were considerable comments and considerable discussion on the guidance document on the parameters that we brought up today.

We certainly have gotten that. It is going to take some time to condense these and bring them together and where we think we need to go with the guidance document. I do not think it needs to be rewritten. I do think it needs to be modified, and I think there is going to have to be some definition in there as to whether or not we can stratify the tumor markers, what basis we stratify them on,

and perhaps some better definition of the analytical parameters that we already have listed for the ones that we are stratifying downward, the clinical information may not be all that critical.

Steve, do you have any comments or any suggestions? Do you want a summary from the panel members?

DR. GUTMAN: No, I think we had a very -- we got our money's worth out of you folks today and we are very grateful. I would like to make a comment that we would like to leave the door open for you to have second thoughts as you are flying home, or for people in the panel to have second thoughts or to share this document or this discussion with their friends and neighbors, and would suggest that anybody who has any brilliant insights over the next 60 days, mail them into Joseph Hackett in our division, and he will be sure to share them with the appropriate people.

DR. REYNOSO: Just to make sure that we somehow continue to consider the concept of parallel studies over time from the same patient as part of the equivalency.

DR. LADOULIS: Any additional questions from members of the public or audience, as now you have heard a number of discussions going on? Any comments anyone would like to raise from the floor? Yes. Step up to the

microphone, please.

MR. FREIBERG: My name is Glen Freiberg, I am with Bard Diagnostic Sciences, I am the Vice President of Regulatory, Clinical and Quality.

As part of the education process for the panel, you have given me an opportunity to give a specific example and also respond to one of the queries that you raised yourself, Dr. Reynoso. In regard to fresh frozen testing, what we had done in the past in our company is we have covered that by getting in fresh samples, freezing alloquots, defrosting them on a weekly basis, and demonstrating that over time, when you defrost them, that you still get the same answer. So that issue, we believe, is covered. Similarly, we do fresh frozen repeats, three or four times, to make sure that the freeze-thaw process does not affect the results we are getting.

That is normally included in our application to FDA, to demonstrate that the sample bank aspect of providing data is valid. In regards --

DR. LADOULIS: For which analytes do you test this?

MR. FREIBERG: For the analytes under review by FDA. Whatever they are, if we are using a sample bank. The

reason I also wanted to bring a specific example is that, in listening to the guidance you provide the FDA, is not fully clear to me that a specific recommendation has been made on the volume of data sometimes when correlation is available.

Just a few weeks ago, working with another partner, we received a letter back from FDA on another PSA test. When this partner came to us and said, we would like to do another PSA test, my first answer was, why? And what I got back was that, if we were to do a high volume micro titre test that was of lower cost to large labs, we believe that there is a market niche for it. It is a tight market niche, but it is a way to go.

We created the antibody and did a ton of data on a correlation aspect from sample banks with the fresh versus frozen, and we provided all those data to FDA. I believe that the data points were at least ten times more than the data Dr. Maxim showed this morning.

In the letter we got back from FDA, or I should say, the sponsor got back from FDA, there is a part of it that says, despite good correlation -- and I would like to stop there. The FDA has said, we have good correlation for a PSA test. Then they go on to say they want additional statistical procedures which I think is not the point I want

to make.

The point I want to make is that after they have stated we have good correlation, this is what was requested. Only single specimens from patients with prostate cancer, and patients with non-prostate cancer malignancies, apparently healthy individuals, and patient with non-malignant diseases, including benign prostatic hyperplasia were examined. There are no data to prove clinical equivalence between any assay approved by the FDA for monitoring disease progression and recurrence.

I am being asked for a whole new demonstration of clinical utility to get this product through, when I have good correlation for another PSA test. For the market I described, for the test I described, this test is dead meat, if the FDA insists on me demonstrating utility again, for a 510(k).

My request to you is to give FDA guidance by saying, if you have good correlation for another PSA test, you do not need that new study.

DR. LADOULIS: Thank you. Any comments from other members of the -- I mean, I can comment but I have talked too much already.

MS. CHACE: Well, I would like to make a comment.

Nina Chace of the FDA staff.

DR. LADOULIS: Yes.

MS. CHACE: And I think maybe to summarize the discussion that we just had about where to put the cut-off, and how good is the analytical sensitivity, specificity? How much do we know about the antigen, how much do we know about the antibody? And where to put the cut-off and that one test you have a different cut-off.

The two tests can correlate, but have you considered where the cut-off is, and if one test could have a cut-off of four and one 75, and still have the same sensitivity and specificity. It seems to me that you have to somehow try to figure out what the sensitivity and specificity is, even if they correlate, because where is the cut-off, and that makes a difference as to what is positive and what is negative.

MR. FREIBERG: That is the problem I have been having is that if I get a three on an approved test and a three on my test; or a five on an approved test and a five on my test -- on the other test. And these tests match up between the predicate and my test. To me, that is substantial equivalence. I do not need to go back and determine with a new patient population that the cut-off is

the same. The correlation is there in all ranges, low, high, cut-off, 'way up, everywhere. You get the same answer from the two tests.

MS. CHACE: But if you do not know the clinical diagnosis of those people --

MR. FREIBERG: Right. So the real question before the panel is, do you believe for another PSA test, that that clinical utility needs to be reproven, and if the answer is yes, it probably should not be a 510(k).

DR. LADOULIS: Dr. Reynoso.

DR. REYNOSO: Yes, I think that we have at least to some degree the panel has provided guidance to the FDA on that issue. I think that -- I may be corrected -- but I think that the consensus as a whole was that there may very well be a group of tests -- do you have an example where analytical equivalency is sufficient?

MR. FREIBERG: Well, this PSA example is the one I am trying to use.

DR. REYNOSO: Well, okay. I am not speaking to any particular test, but to the idea. To the idea, but there is a number of tests or a number of situations, where analytical equivalency, correlation, may be sufficient, and no additional clinical studies are needed. And my

interpretation of what we said today is that the FDA is going to try to modify the guidelines to reflect some of these distinctions that have been made this afternoon. Am I correct?

DR. MAXIM: That is exactly correct. That was the whole purpose of the meeting today. I think on several different occasions we stated that, although we uniformly applied the criteria of the guidance document across all assays, that it may not be necessary in all cases, and that is why we are getting panel input and comment today to make those modifications. I thank both of you for your comments.

DR. TAUBE: And we also said clinical equivalence, rather than clinical utility, and I think there is a difference.

MS. CHACE: Could I ask Mr. Freiberg, if you did the normal range study and were the two normal --

DR. MAXIM: I do not think that is appropriate for this forum.

MR. FREIBERG: The question Dr. Taube asked I am not sure I know how to answer, because I do not know how I would do a clinical study to separate the two concepts.

DR. LADOULIS: No, I think she was just restating what you said.

DR. TAUBE: No, what I was restating was what the panel has said today.

DR. LADOULIS: That is the clinical --

DR. TAUBE: Which is, there is a difference -- you do not have to do a whole clinical utility study showing every possible outcome, but you have to show clinical equivalence. That is, a patient who has disease. And it is not just a number of three or four, but a patient who has disease with one test has disease with the other test, because that is where the cut point --

MR. FREIBERG: But all the bank samples are well-characterized patients, we know -- we have all their clinical data.

DR. TAUBE: Well, we cannot discuss the specifics of your submission.

DR. LADOULIS: But if in performance it is equivalent for single point determination, I think that is what we were discussing.

MR. FREIBERG: Thank you for the opportunity --

DR. LADOULIS: -- performance characteristics are there, that longitudinal studies may not be required, except as Dr. Reynoso referred to. So, thanks for the comments. I think we have covered that issue, and are there any other

questions that you wanted to bring to the panel?

DR. MAXIM: I simply want to thank the panel again for coming together today to discuss these issues, and thank you for your valuable input on behalf of the Center. Once again, we thank you for your cooperation and participation in the review process.

DR. LADOULIS: If that is the end, if there are no others, then this is adjourned.

[Whereupon, at 3:08 p.m., the meeting was concluded.]